TECHNICAL REPORT

APPROVED: 28 June 2019

doi:10.2903/sp.efsa.2019.EN-1673

# Outcome of the Pesticides Peer Review Meeting on general recurring issues in ecotoxicology

**European Food Safety Authority**

Maria Arena, Domenica Auteri, Stefania Barmaz, Eugenia Chaideftou, Lucie Ctverackova, Chloe De Lentdecker, Alessio Ippolito, Dimitra Kardassi, Chris Lythgo, Tunde Molnar, Laura Padovani, Rachel Sharp, Franz Streissl, Juergen Sturma, Csaba Szentes, Benedicte Vagenende, Joanke Van Dijk and Laura Villamar-Bouza

## Abstract

This technical report reflects the outcome of the ecotoxicology experts' meeting on general recurring issues noted during the EFSA peer reviews of pesticide active substances under Regulation (EC) No 1107/2009. General and specific issues were identified and discussed relating to risk assessment for birds and mammals, aquatic organisms, non-target arthropods and soil organisms. Conclusions and recommendations on these topics were drawn.



**Key words:** risk assessment, birds and mammals, aquatic organisms, non-target arthropods and soil organisms

**Requestor:** EFSA

**Question number:** EFSA-Q-2019-00344

**Correspondence:** pesticides.peerreview@efsa.europa.eu

**Acknowledgements:** EFSA wishes to thank the following experts for the support provided to this scientific output: McVey E, Alonso N, Gioutlakis M, Duquesne S, Pedersen S, Marchetto F, Fryer M, Kraemer W.

**Suggested citation:** EFSA (European Food Safety Authority), 2019. Technical report on the outcome of the Pesticides Peer Review Meeting on general recurring issues in ecotoxicology. EFSA supporting publication 2019:EN-1673. 117 pp. doi:10.2903/sp.efsa.2019.EN-1673
**ISSN:** 2397-8325

## Summary

During the EFSA peer review of pesticide active substances under Regulation (EC) No 1107/2009, several issues in the area of ecotoxicology were identified by EFSA and Member States that needed discussion with experts from national authorities in order to enhance the harmonisation of the risk assessment of active substances.

General and specific issues related to risk assessment for birds and mammals, aquatic organisms, non-target arthropods and soil organisms were identified and discussed in a general ecotoxicology meeting, Pesticide Peer Review Meeting 185, which took place from 9 to 12 October 2018.

Recommendations on these topics were compiled based on the discussion and conclusions achieved at the meeting. These recommendations will be applied during the EFSA peer review of active substances and they are expected to provide additional clarifications to applicants and rapporteur Member States regarding the scientific interpretation of the relevant issues when preparing the dossiers and the assessment reports. Furthermore, it is expected that these recommendations will be taken into consideration during the revision of the relevant guidance documents in the area of ecotoxicology.

# Table of contents

# 1. Introduction

During the EFSA peer review of pesticide active substances under Regulation (EC) No 1107/2009[1], EFSA identified several general recurrent issues in the area of ecotoxicology which warranted expert consultation and agreement in order to enhance the harmonisation of the risk assessment process for active substances.

To this purpose a second general meeting was organised and took place in October 2018 (Pesticide Peer Review Meeting 185, 9–12 October 2018). Representatives with expertise in ecotoxicology from 20 Member States attended this meeting, with good coverage across the southern, central and northern zones (see Figure 1).

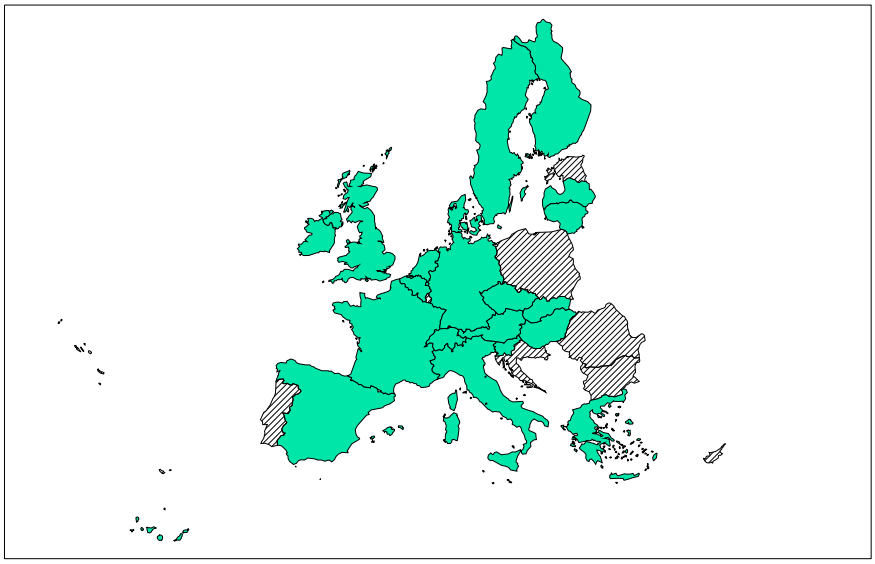


**Figure 1**: In green the countries that were represented at the meeting, i.e. 20 Member States plus Switzerland as an observer

In addition to the points identified by EFSA during the peer review of pesticide active substances, Member States were requested to collect and submit to EFSA issues which are relevant at zonal level. The final agenda of the meeting was developed by considering both EFSA and Member States' proposals.

The issues identified related to both general and more specific points in the area of risk assessment for birds and mammals, aquatic organisms, non-target arthropods (NTAs), soil organisms and terrestrial non-target plants. They are described in the following sections.

In addition, the following documents are available as background documents to this technical report:

- the report of the meeting;
- the comments received on the draft technical report following the written procedure conducted from 1 to 22 April 2019. It is noted that the written procedure was performed with the purpose of enhancing readability and to correct possible inconsistencies. Since the scope of this technical report was to reflect the meeting discussions and conclusions, the commenting round was not meant to reopen the discussions or to change the outcome of the meeting.

[1] Regulation (EC) No 1107/2009 of the European Parliament and of the Council of 21 October 2009 concerning the placing of plant protection products on the market and repealing Council Directives 79/117/EEC and 91/414/EEC. OJ L 309, 24.11.2009, p. 1–50. Available online: http://data.europa.eu/eli/reg/2009/1107/oj

# 2. General issues

## 2.1. Extrapolation of studies between different agroclimatic conditions

This issue was proposed and presented by Member States of the southern Europe zone, according to the EU zones as defined by Regulation (EC) No 1107/2009. Concerns were raised regarding the representativeness and appropriateness of extrapolating higher tier data between zones for mesocosm studies conducted in the northern zone and for higher tier studies on birds and mammals (e.g. the identification of focal species; the composition of the diet of these species obtained from the treated area; the residue levels of pesticide in this diet; the proportion of the daily diet that focal species obtained from the treated area (PT)).

In the case of mesocosms, the majority of the experts at the meeting agreed that the no observable effect concentration (NOEC) and the ecological threshold option (ETO) regulatory acceptable concentration (RAC) can be used in the risk assessment with the assessment factor (AF) recommended by aquatic guidance (EFSA PPR Panel, 2013), and this can be considered as independent of the experimental conditions (e.g. the climatic zone). However, when an ecological recovery option (ERO) RAC is derived, the extrapolation between zones should be considered carefully taking into account the fact that the ability for recovery may vary pending on the agroclimatic conditions. A case-by-case evaluation should be carried out, based on the information available. For further consideration on what is covered by the AF, see Section 4.3.

In relation to the higher tier studies for birds and mammals, the experts considered that the recommendations given by EFSA (2009) are sufficient for spray applications, i.e. any refinements of the risk based on identification of specific focal species and definition of related ecological data should be representative of the area of use of the active substance. This means, for example, to extrapolate a focal species from one zone to another requires consideration of whether the criteria for selecting the focal species are still met. However, the experts noted that higher tier studies for seed treatment uses would need further attention, in order to take into account specific agronomic practices (e.g. sowing rates) and conditions. The experts suggested that any issue related to the agronomic practices may be addressed in the European Commission's guidance document on seed treatments which is under development and can be considered in the context of the revision of the EFSA Guidance (EFSA, 2009).

## 2.2. How to consider studies when the analytical methods are not validated

In line with Commission Regulation (EU) No 283/2013[2], methods for the determination of non-isotope-labelled residues used in support of ecotoxicology studies should be generated and reported in the dossier. This information should be provided both for old studies (of the original peer review) and new studies (for the renewal). This is applicable to all areas of the risk assessment (i.e. for the purposes of testing toxicological, ecotoxicological, environmental, residue and physico-chemical properties). The usual matrices of interest in the case of the ecotoxicity testing are soil, water, sediment and feedstuffs (European Commission, 2000).

Currently, the validation of the analytical methods is performed in the physico-chemical properties area and the related assessment is reported in Volume 3, Chapter B.5. When methods are not fully validated, the experts responsible for the other sections should be informed (see EFSA (2017a) for further details).

It is noted that, mostly in the case of approval for the renewal of active substances, often the methods in the 'old studies' (e.g. those performed before the publication of Regulation 283/2013), cannot be validated in accordance with the current guidance (European Commission, 2000). In those cases, depending on the available information and on the basis of the expert judgement, it could be

[2] Commission Regulation (EU) No 283/2013 of 1 March 2013 setting out the data requirements for active substances, in accordance with Regulation (EC) No 1107/2009 of the European Parliament and of the Council concerning the placing of plant protection products on the market. OJ L 93, 3.4.2013, p. 1–84.

concluded that a method is not validated but nevertheless is fit for purpose and, therefore, supports the ecotoxicity studies.

To enhance the harmonisation of the evaluation of this issue in the assessment reports, it was considered and discussed that the validation status of the analytical methods should be considered in the appraisal of the quality of each ecotoxicity study. The validity of the studies for which the analytical methods are not validated or considered fit for purpose should be questioned. However, for the sake of reducing the vertebrate testing, the repetition of a study on vertebrates should be carefully considered. This approach is also followed for mammalian toxicology studies (EFSA, 2016).

The experts at the meeting agreed that where the method is not validated or not fit for purpose, a case-by-case evaluation should be conducted. All the available information, including the toxicological profile of the substance and the margin of safety of the risk assessment, should be considered before rejecting studies. The applicants should be requested to provide justifications to support endpoints from studies where the analytical method was not fit for purpose. In the event that a study supported by a method not fit for purpose is used in the risk assessment this should be flagged in the list of endpoints.

Additionally, it was recommended that in Volume 3 Chapter B.9 of the renewal assessment reports (RARs) the conclusion of the assessment on the validation the analytical method should always be reflected as part of the evaluation of each ecotoxicological study. In line with previous agreements (EFSA, 2017a), the related assessment should be reported in Volume 3 Chapter B.5.

Some examples of fit-for-purpose analytical methods are given in Appendix A.

## 2.3. Risk assessment for plant protection products

### 2.3.1. How to consider the formulation within the evaluation of the active substance

Regulations (EU) 283/2013 and 284/2013[3] set out the data requirements for active substances and plant protection products (PPP), respectively, (including requirements for ecotoxicological data for both the active substances and the PPP).

According to Regulation (EU) 283/2013, Section 8, for the approval of the active substance, data not only on the active substance but also on the PPP might be submitted, depending on which information is more appropriate to address the toxicity. This is reported as follows:

'In the case of certain study types, the use of a representative plant protection product instead of the active substance as manufactured may be more appropriate, for example testing of non-target arthropods, bees, earthworm reproduction, soil micro-flora and non-target terrestrial plants. In the case of certain plant protection product types (for example encapsulated suspension) testing with the plant protection product is more appropriate to testing with active substance when these organisms will be exposed to the plant protection product itself. For plant protection products where the active substance is always intended to be used together with a safener and/or synergist and/or in conjunction with other active substances, plant protection products containing these additional substances shall be used.'

According to Regulation (EU) 284/2013, when the toxicity cannot be predicted from the active substance or when the results of the acute toxicity study indicate higher toxicity of the formulation, studies performed with the PPP are required. This means that the standard assessment presented for the active substance will not be sufficient to conclude on the risk from both active substance and formulation and specific studies would be performed on the PPP. This is mentioned in several places and in the specific sections in the Regulation.

The purpose of this discussion point was to achieve a better understanding and enhance the harmonisation between Member States on how to consider the toxicity of the formulation relative to the toxicity of the active substance and how to deal with the risk assessment of the PPP within the

[3] Commission Regulation (EU) No 284/2013 of 1 March 2013 setting out the data requirements for plant protection products, in accordance with Regulation (EC) No 1107/2009 of the European Parliament and of the Council concerning the placing of plant protection products on the market. OJ L 93, 3.4.2013, p. 85–152.

peer review of the active substances. The discussion concerned those situations in which some data on both the active substance and formulation are available in the EU dossier (usually only for acute toxicity). In particular, EFSA proposed for discussion two main points for the different groups of non-target organisms:

- In which situations should a formulation be considered as being more toxic than the substance under assessment?
- What is the best approach to take when a formulation is more toxic and a comprehensive risk assessment has not been performed?

In relation to 'when a formulation should be considered more toxic than the active substance', the proposal was to account for a difference of a factor of three, as recommended in the guidance from the Directorate-General for Health and Food Safety (SANCO/10597/2003 rev. 10.1) (European Commission, 2012) on the equivalence of batches and in the aquatic guidance (EFSA PPR Panel, 2013). This means that when the endpoint of the PPP (expressed in terms of the active substance) is at least three times lower than the equivalent endpoint for the active substance, it should be considered to be more toxic. This factor was agreed by the majority of the experts, to be applied consistently to Tier 1 studies for all groups of non-target organisms.

For birds and mammals, the data on mammals from the mammalian toxicology section should be considered first. If, based on the comparison of data on mammals, it is clear that the formulation is more toxic, it was agreed that the risk assessment should be performed based on the formulation endpoint, expressed in terms of the active substance, as reported in Regulation (EU) 284/2013. However, before asking for further vertebrate studies (e.g. on birds), other elements should be considered, such as the margin of safety in the risk assessment for mammals or factors which may have an impact on the overall toxicity of the formulation (e.g. carriers, dose spacing, method of dosing).

In the case that multiple studies are available that give contradictory information in terms of the comparison of toxicity between active substance and formulation, it was recommended that all the available data should be considered and a decision made on a case-by-case basis; for example, by considering the sensitivity of the tested species.

For aquatic organisms, if the formulation is more toxic than the active substance, the majority of the experts considered that separate risk assessments for the active substance and for the formulation with their respective endpoints could be provided. In the absence of a comprehensive exposure characterisation for the formulation, the predicted environmental concentrations in surface water ($PEC_{SW}$) values generated for the active substance accounting for all the routes of exposure should be used in combination with the formulation endpoint expressed as active substance.

For bees and soil organisms, if the formulation is more toxic than the active substance, the majority of the experts agreed to follow the same approach as described above for the aquatics, i.e. to perform separate risk assessments: one with the active substance and the other with the endpoint for the formulation expressed as active substance.

Some experts expressed the concern that when more than one substance is included in the formulation, the approach of assuming that the toxicity is entirely due to the substance under evaluation may result in a too conservative risk assessment. This is because the entire toxicity of the formulation will be attributed to the substance under evaluation. However, the approach agreed at the meeting is in line with Regulation (EU) 284/2013 and will only be used when an applicant does not provide a comprehensive formulation risk assessment.

There was no discussion on this point for NTAs and non-target terrestrial plants, since only data on formulation are usually available for these organisms. Where data on the active substance and on the formulation are available, a separate risk assessment should be performed as for the other organism groups.

Overall, it can be concluded that when a PPP appears to be more toxic, i.e. its toxicity endpoint is three times lower than the equivalent endpoint of the active substance, according to the data requirement the lower endpoint should be used for the risk assessment or risk assessments for both the active substance and PPP could be provided.

### 2.3.2. Tests with formulations containing more active substances with different degradation times

The issue proposed related to the evaluation and the expression of endpoints from aquatic toxicity studies performed with formulations containing more than one active substance with different degradation times. In such studies, analytical measurements of each active substance should be performed at test initiation and test termination (fish, invertebrates and algae). For chronic studies intermediate time point measurements are usually performed.

It was noted that in some cases the chemical analysis is only performed on one of the active substances. In other cases, chemical analysis is performed for all active substances within the product but the concentration of one of the substances is not adequately maintained during the study. The discussion among the experts aimed to achieve a harmonised approach on how to consider those studies and on how to express the endpoints.

Overall, the experts agreed that, unless it is clear which substance drives the toxicity, all the active substances in the formulations should be measured and the stability should be confirmed. If the concentration of one of the active substances is not maintained during the study, it should not be considered as a Tier 1 study but it may still be used for Tier 2 risk assessment (Section 2.3.3 of this technical report maybe consulted for those cases).

### 2.3.3. Feedback from central zone harmonisation meeting

A proposal regarding issues with formulation testing in relation to the risk assessment for aquatic organisms was introduced by the Member States of the central zone. The issue related to the expression of endpoints from Tier 1 tests and formulation tests (with one or more active substances) for unstable substances. The proposal was in a draft phase at the time of the meeting and was therefore not discussed. After the meeting, the final version of this proposal agreed by the Member States from the central zone was shared with EFSA. This proposal is included in Appendix J because it reports recommendations which could be useful for other Member States conducting risk assessments for the authorisation of PPPs. Also, this proposal may be consulted by rapporteur Member States when preparing risk assessments for active substance approval (see also Section 2.3.2).

## 2.4. Use of residue data to support ecotoxicological assessments

The ecotoxicological risk assessment of pesticides is increasing in complexity where input data from other disciplines (e.g. environmental fate and behaviour) need to be incorporated for a more appropriate characterisation and quantification of the exposure for non-target taxa.

Information in the residue section (Volume 3 Chapter B.7. of the draft assessment report (DAR)/renewal assessment report (RAR)) aims to describe and quantify the residues that may enter the consumers' diet. However, this information, together with the data provided by environmental fate specialists, might support the identification of residues that may need further consideration for the dietary risk assessment of non-target organisms (e.g. birds and mammals and bees).

The studies listed below are required under the data requirements for PPPs (Regulation 283/2013):

i. Metabolism studies in primary crops
ii. Metabolism studies in rotational crops
iii. Supervised residue trials
iv. Metabolism studies in livestock
v. Feeding studies
vi. Other studies (residues in pollen and bee products).

A more detailed description of the studies listed above is given in Appendix B, as well as certain considerations for the interpretations of the results.

The point proposed for discussion during the general meeting was related to how to integrate information contained in the residue section (DAR/RAR Vol. 3, B.7.) in order to inform the environmental risk assessors.

There was a general consensus at the meeting that there is a need to establish an efficient and consistent communication between the residue and the ecotoxicological sections. As agreed with the experts, EFSA has developed a questionnaire to consistently collect and report the relevant information. An example of how to fill in this questionnaire is included in Appendix C. At the end of that appendix, a link is provided to download a blank template that can be used at Member State level and reported as an Appendix to Vol. 3, B.9 as supporting information, if desirable.

## 2.5. Equivalence of batches

The issues proposed for discussion were:

1) Whether the concentrations and subsequent endpoints should be corrected for the purity of the test item. This is primarily relevant for studies where chemical analysis is not routinely performed or when the endpoint is expressed in terms of nominal concentration.
2) To agree on the best way to present and conclude on the equivalence of the batches used in the ecotoxicity studies.

In relation to point 1, the experts at the meeting agreed that for substances with less than 90 % purity, when the endpoints are expressed in terms of nominal concentrations, these should be corrected for the purity of the technical material. It must be noted that in such situations the tested item is to be considered a mixture. Expressing the endpoint in terms of pure active ingredient content may overestimate the toxicity of the active substance, but it would ensure consistency when the toxicological endpoint is compared with the exposure estimates in the risk assessment.

In relation to point 2, the experts agreed to report in Vol.3 B.9 of the assessment reports studies for which the compliance of batches was not demonstrated. As agreed at the meeting, a template for how the assessment of the compliance of the batches with the technical specification (new and old, if any) should be reported in Volume 4 has been developed and included in Appendix D. It was agreed that an overview of the batches used in all the available ecotoxicological studies should be presented in line with the Commission guidance (European Commission, 2012): a Tier 1 assessment should be presented for all the batches used in the ecotoxicological studies while a Tier 2 assessment should only be performed for those batches used in key studies (i.e. studies used for risk assessment).

Studies using batches which have not been demonstrated to be equivalent to the technical material should also be flagged in Volume 3. There was a consensus that, in general, the issue is not of such significance to identify a critical area of concern and only a data gap should be identified in the EFSA conclusions in situations where it has not been demonstrated that the material in the ecotoxicity studies complies with the technical specifications. However, where the available information indicates a potential concern (e.g. impurity considerably more toxic than the active substance), then a critical area of concern may be identified in the EFSA conclusion.

## 2.6. Use of $EC_{10}$ values in environmental risk assessments

In the first general ecotoxicology meeting (Pesticides Peer Review Meeting 133) the evaluation of the reliability of $EC_{10}$ calculations were discussed and some guidance was developed, as reported in the technical report of the meeting (EFSA, 2015). A follow-up discussion was proposed for the second general meeting, in order to consolidate the previous agreement.

The experts at the meeting concluded that an update of the guidance given in Appendix F of the technical report (EFSA, 2015) was needed. The update is included in Appendix E of this report, which gives a synthesis of the whole process and the agreed approach.

## 2.7. Risk assessment for rice paddies

This point was proposed for discussion by Italy, which is leading the development of a specific guidance for environmental risk assessment for the southern zone. The following background was presented at the experts' meeting.

Rice is the only European crop cultivated in fields regularly flooded for long periods. This means that rice fields can be considered water environments, similar to swamps, ponds and other moist environments. However, guidance documents are developed for risk assessments of PPPs and active

substances on the basis of application to 'classic' crops that represent terrestrial ecosystems. The intrinsic diversity between the two kinds of environment leads to discrepancies in risk characterisation. Moreover, the huge amount of water required by rice cultivation leads to the development of complex systems of channels and ditches to bring water in and out of the paddies, which in their turn constitute closed and waterproof chambers where rice plants can be submerged during their growth. The systems of channels and paddies often constitute a continuum of water bodies separated by thin banks highly managed by farmers. Consequently, it was suggested that the concept of 'off field' in rice system cultivation is different from that for other crops.

Currently, there is only one guidance document dedicated to the environmental risk assessment of active substances used on rice (European Commission, 2003), which was developed by an ad hoc working group. This document, however, simply suggests applying the other guidance documents in force at the time of its development. Member States that are rice producers organised a task force to draft ad hoc guidance to address the approval of active substances and the authorisation of PPPs proposed for application on rice, reviewing the current guidance in order to adapt the current scenarios or to fill the gaps that arise from the differences related to this peculiar environment.

Italy, the leader of the task force, presented the current ongoing activities. The experts welcomed the work being undertaken by Italy and the task force and agreed that rice should not be overlooked in the development of guidance documents. It was noted that specific recommendations for rice crops should be developed as part of the standard guidance documents or as specific guidance.

## 2.8. Risk assessment for banana crops

This point was proposed for discussion by Spain and the following background and proposals were presented at the experts' meeting.

The largest producer of bananas in the EU is Spain (60%) in the Canary Islands, followed by France (36%), in Martinique and Guadalupe, and Portugal (3%) in Madeira. The Canary Islands are a group of seven islands off the Moroccan coast of Africa with Tenerife and La Palma as the most important islands for banana production. Therefore, Tenerife and La Palma were suggested as representative locations. Banana-growing in Tenerife is located almost exclusively in the coastal strips on the northern and western sides of the island. Due to the geographical location of the Canary Islands, some 4º from the tropic of cancer and very close to the African coast, the islands' climate is subtropical.

It is noted that the banana is a giant herbaceous plant that can be harvested throughout the year. Bananas can be harvested four months after planting, which can result in up to three harvests in a year.

Regarding the exposure assessments for soil and surface water, there is no specific scenario for banana crops. However, the Southern Zone Guidance (South Member States, 2018) provides specific recommendations for exposure calculations for banana crops.

Since, in the currently available guidance, scenarios for bananas are not available, the risk assessment for non-target organisms is usually conducted with orchard as a surrogate crop.

The biodiversity of the Canary Islands is conditioned by its subtropical climate, its volcanic origin and its location. Risk assessments for birds and mammals should be conducted with specific focal species which are relevant to this area. In addition, it was proposed that, since the Canary Islands have a great abundance and variety of reptile species, focal species for this group of vertebrates should also be defined.

Banana crops can be grown near to the coast, mainly on Tenerife Island. Since marine species can be more susceptible to pesticides than freshwater species, it was proposed that marine species should be considered for aquatic risk assessment.

Regarding risk assessment for banana crops, a discussion took place following a proposal by Spain on:

- Defining the characteristic scenarios for banana crop, considering the specific agroclimatic conditions of the Canary Islands.
- Risk assessment should be conducted with focal species traits from the Canary Islands.

- Higher tier studies should be performed with relevant species.

The experts at the meeting noted that the uncertainty regarding the risk to marine species is not specific to banana cultivation and further analysis of data should be carried out to understand whether marine species are in fact more sensitive than freshwater species, e.g. by using data from US dossiers or from biocides. The experts also noted that the identification of appropriate focal species for birds and mammals may be challenging given the high variability of the geographic location of EU banana plantations, which are not exclusively the Canary Islands. Nevertheless, according to EFSA (2009), specific focal species used for higher tier risk assessment must be relevant to the area of use. The experts noted that the lack of agreed assessment methodology, including the identification of appropriate focal species, for the risk assessment of reptiles is not specific to bananas. Overall, it was recognised that some aspects, such as the assessment of risk to terrestrial vertebrates, might be further considered in the revision of the EFSA Guidance (2009) and in the ongoing statement for bat species from the EFSA PPR Panel. The orchard scenario in the EFSA Guidance (2009), was considered to be a reasonable interim solution for Tier 1 risk assessment for bird and mammal risk assessment for bananas. In the event that a higher tier assessment is triggered, the applicant(s) must provide suitable information to identify relevant focal species and ecological information relevant to the specific location, if used for refinement.

# 3. Birds and mammals

## 3.1. Trials for residue decline

This issue was proposed for discussion in order to achieve harmonisation of the consideration of field residue decline studies used to refine the dissipation in plant material and invertebrates in the context of the assessment of risk to birds and mammals. In particular, the following aspects were proposed for discussion and agreement:

- criteria to assess the reliability of single studies (field trials)
- criteria for extrapolation within and between items (matrices)
- dissipation kinetics of a single trial
- how to deal with multiple applications
- use of degradation kinetics in the risk assessment (e.g. minimum number of trials)
- merging data sets
- metabolites.

The points above are presented in detail in Appendix F.

The experts and EFSA highlighted that this issue would be better addressed by the working group undertaking the revision of EFSA (2009). The elements considered in this technical report could form the basis for the discussion in the working group, and, where relevant, provide an interim solution in the context of the assessment of risk to birds and mammals according to EFSA (2009).

The suggested criteria for assessing the reliability of single studies (field trials) were discussed and agreed for plant materials and invertebrates. They are related to aspects of the experiment design and the reporting for the field and the analytical phase. The agreements and conclusions were consolidated and are reported in Appendix F.

In relation to extrapolation within and between items, EFSA has recommended several broad groups where extrapolation may be considered acceptable:

- Dicot plants (green parts and roots)
- Monocot plants (green parts and roots)
- Fruits
- Seeds (both weed seeds and cereal seeds)
- Foliar-dwelling arthropods

- Ground arthropods
- Earthworms.

Overall, the experts at the meeting agreed with the extrapolations as suggested by EFSA. However, it was pointed out that generally common sense and expert judgement should be used, for example, for crop extrapolation within monocot and dicot groups. Extrapolation between the above groups was not considered appropriate. In case of dicot weeds and grass-like weeds, extrapolation is possible for trials performed at a late growth stage. It was also agreed, in general, to avoid extrapolating from maize to grass-like weeds, because maize is a fast-growing crop.

As regards the dissipation kinetics of a single trial, the recommendations given by FOCUS (2006) were reported. In general, for plants, the single first-order (SFO) kinetic model was recommended. It was agreed that a minimum of five time points should be available for fitting. However, in some exceptional cases four points may be enough (e.g. fast dissipation or metabolites) but there should never be fewer than four time points. Some experts highlighted that the use of pseudo $DT_{50}$ obtained with the first-order multi-compartment (FOMC) model may be appropriate when SFO cannot be used. For invertebrates, in the case that an attempt to fit a kinetic model provides unreliable $DT_{50}$ or pseudo $DT_{50}$, a proposal could be to calculate the time-weighted average factor ($f_{TWA}$) by integrating the area under the curve (AUC) normalised by the initial value and divided by the averaging period (generally 21 days). The use of the AUC directly in the risk assessment would mean ignoring the residue per unit dose database given in the EFSA Guidance and this is not recommended. Hence, this kind of approach should only be used to derive an $f_{TWA}$.

In relation to the use of degradation kinetics in the risk assessment (e.g. minimum number of trials), the EFSA proposal for plant materials (residue trials performed at least four sites per item and regulatory zone) was agreed although some uncertainties were pointed out. However, it was also agreed that particular climatic conditions of certain areas should be considered to allow extrapolation to some extent (e.g. northern France).

For invertebrates, an agreement between the experts regarding the minimum number of trials or sites could not be achieved and it was considered as an issue to be dealt with by the ongoing working group for the revision of the EFSA Guidance (2009).

The proposed discussion points related to 'dealing with multiple applications', 'merging data sets', and 'metabolites' were not discussed at the meeting. The EFSA proposals, as reported in Appendix F were commented on by the Member State experts during the written procedure. Overall, the proposal was generally accepted by Member States as an interim solution.

## 3.2. 21-day PT

In the context of higher tier bird and mammal risk assessment, one important issue is the consideration of the proportion of the total food intake of an animal that is obtained from consuming food in a pesticide-treated area (PT). The initial, worst-case assumption is that 100% of food is obtained from the treated area. One of the potential refinements is to seek, and use, a more accurate measure of the fraction of the total food intake that an individual animal obtains from a pesticide-treated field. It is usual for this refinement to be based on radio-tracking individuals of a species of particular concern or relevance (i.e. a 'focal species').

Most radio-tracking studies have comprised an observation period of one day, or several shorter-term observation periods being brought together to create 'a day in the life' of a particular bird or mammal. However, such PT data are often introduced into long-term/reproductive risk assessments, i.e. where birds or mammals might have received exposures spread over the period of their reproductive cycle. Hence there is a potential mismatch between the time period used to assess the PT and the characteristics of the risk assessment being undertaken.

Recently, Ludwigs et al. (2017) published work on how to produce a 21-day time-weighted average value for PT for the wood pigeon. The rationale behind this work was to try to better match the potential PT refinement step to the needs of the long-term/reproductive risk assessment. The methodology involves using a Monte Carlo approach. It should be noted that the aim of the paper was to highlight the potential use of the methodology rather than produce a specific refined PT value for wood pigeon. The methodology developed has the potential to be used in risk assessment by reducing

the PT value from that conventionally arising from the 'day in the life' approach. In light of this, the UK's Health and Safety Executive evaluated the paper and raised several concerns and suggested potential alternatives.

The UK argumentations on this paper were discussed at the meeting for agreement between Member States. In general, the potential of the approach described in Ludwigs et al. (2017) was recognised, but it was agreed that further consideration on how to use it for risk assessment purposes would be needed. It was further noted that the current approach is based on consumers and the number of consumers is likely to increase with time but the PT for the individual could decrease with time. Therefore, the approach may become less conservative than taking single-day values. The experts recommended that the working group on birds and mammals should reflect on this. For the time being, it was agreed not to use this approach.

# 4. Aquatic organisms

## 4.1. Use of geometric mean and weight of evidence for acute data

When data are available on multiple species, but are insufficient to carry out a species sensitivity distribution (SSD), the aquatic guidance document (EFSA PPR Panel, 2013) foresees the possibility for combining the endpoints in a geometric mean, to be used in the risk assessment together with the standard Tier 1 AF (100 for acute and 10 for chronic). However, the same guidance highlights that this approach may be problematic when the geometric mean is biased by the introduction of insensitive species. In order to avoid using an RAC which is not sufficiently protective, the guidance recommends using the $RAC_{geomean}$ only when this is lower than the lowest available endpoint. When this is not the case, the guidance suggests adopting a weight-of-evidence approach, i.e. to use the lowest available endpoint with a reduced AF. However, no further specific indications are given therein.

At the meeting, some experts highlighted that the geomean approach generally provides an uncertain level of protection. Therefore, a decision scheme more elaborated than the existing recommendations from the (EFSA PPR Panel, 2013) had been proposed in order to account for all possible scenarios.

According to this proposal, every time data on multiple species are available, two separate RACs should be calculated: the standard $RAC_{geomean}$ and another RAC based on the lowest endpoint with a modified AF ($RAC_{lowest}$). Two cases were discussed.

**Case 1** represents the situations when the $RAC_{geomean}$ > lowest endpoint, and the proposal would be to have a modified AF ≥ 20 for the $RAC_{lowest}$. **Case 2** represents the opposite case, and the proposal would be to have a modified AF ≥ 60 for the $RAC_{lowest}$.

In addition, for the risk assessment, it was proposed to use the following RAC:

- For case 1, the $RAC_{lowest}$ (lowest endpoint and AF ≥ 20)

- For case 2:

  - the $RAC_{geomean}$ for active substance assessment (flagging any case where $RAC_{geomean}$ > $RAC_{lowest}$)
  - the lowest of the two RACs for the product authorisation.

All experts considered that **case 1** is somehow already contemplated in the (EFSA PPR Panel, 2013), and all were willing to immediately implement a harmonised approach. After considering the historical components of the AF for acute data (10 for covering intra- or inter-species variability and 10 for covering other uncertainty aspects), the maximum AF used for the SSD and the minimum number of species required for the SSD, some experts proposed that the modified AF should be ≥ 20 for invertebrates and ≥ 30 for fish. This proposal is based on the assumption that the AF can be reduced linearly by increasing the number of tested species. However, this assumption should be further checked in the future. It was also noted that, not only the number of species, but also other elements should be considered when establishing the modified AF (e.g. how much of the taxonomic diversity is covered, the reliability of the available studies, etc.). Overall, all experts agreed on the proposed approach for case 1, setting the minimum modified AF to 20 for invertebrates and to 30 for fish.

However, several concerns were raised at the meeting regarding the approach for **case 2**. Some experts disagreed with having different approaches at the EU and at zonal/country level. Regarding this point, however, some interim solutions were proposed, in order to guarantee a certain level of harmonisation (e.g. introduction of standard text flagging the potential lack of protectiveness when the $RAC_{geomean}$ is only slightly higher than the lowest available endpoint). It was noted that this approach has already been applied in the northern zone and several Member States in the central zone would be willing to implement it immediately. However, several other experts considered that such a proposal goes beyond what is currently suggested by the (EFSA PPR Panel, 2013) and were therefore reluctant to implement it before a proper calibration has been performed. Therefore, no agreement could be achieved for the approach to be taken for case 2.

All experts agreed that the whole approach (covering both case 1 and case 2) should be further considered in the context of the revision of the (EFSA PPR Panel, 2013).

## 4.2. Use of geometric mean and weight of evidence for chronic data

A similar approach to that described in Section 4.1 (for acute data) was also proposed for chronic data. Such a proposal would be limited to $EC_{10}$ (animals) and $EC_{50}$ (primary producers). Additional considerations were included for use with vertebrates. Some experts showed sympathy for the proposal, either for the full approach or just for agreeing on a strategy for lowering the AF when data on multiple species are available.

However, at the meeting it was highlighted that the use of the geometric mean for combining chronic data is currently not supported, due to concerns raised by some Member States after the publication of the AGD, which contemplates this approach. Some experts reported that in their countries the use of the geometric mean for chronic data is limited to algae and aquatic plants, while for vertebrates and invertebrates, discussions are still ongoing.

Overall, the experts agreed to maintain the status quo (i.e. not to support the use of the geometric mean for chronic data, nor an arbitrary reduction of the AF in a weight-of-evidence approach), but all expressed the wish to reconsider the issue in the next revision of the (EFSA PPR Panel, 2013).

## 4.3. General recommendations on mesocosm experiments

While acknowledging that the part on model ecosystems (micro-/mesocosms) as included in the aquatic guidance (EFSA PPR Panel, 2013, Section 9.3) is considered to be very helpful, some Member States felt the need for some aspects to be further clarified, as they are not always appropriately addressed by applicants.

At the meeting, there was a general agreement on the recommendations presented, which are therefore reflected in this technical report (below). All of the following issues should be considered in the next revision of the (EFSA PPR Panel, 2013).

**Representativeness and vulnerability of the communities tested**

The AF applied to the NOEC or NOAEC (for deriving the ETO or ERO RAC) is used for spatio-temporal extrapolations (for values of the AF, see (EFSA PPR Panel, 2013) p. 127; tables 34 and 35); it does not cover other elements (e.g. low representation of some vulnerable taxa).

It should be considered that the community represented is usually dominated by R-strategists, with high reproductive potential, and which are therefore of low vulnerability. This concern is particularly relevant for ERO derivation.

For invertebrates, this concern can be addressed by ensuring a sufficient number of EPT (Ephemeroptera, Plecoptera, Trichoptera) species. These taxa are generally quite vulnerable due to their reproductive cycles and to their high sensitivity to some substances. It is noted that EPT are also an important component of a functioning ecosystem[4]. At the meeting, it was, however, noted that

[4] EPT are used under the EU Water Framework Directive (Directive 2000/60/EC) to assess the health of numerous types of water bodies in the EU, according to the Lenat index (1988), which links the proportion and types of EPT present to the overall health of the aquatic ecosystem.

these taxa are generally not particularly abundant in mesocosms, and that most of them prefer cold fast-running water, while most mesocosm experiments are carried out in pond-like structures. Some experts also suggested that it may be appropriate to build up a list of the species/taxa which should be present in the mesocosms.

On a practical level, an absence or low abundance of these groups should not necessarily result in the invalidation of the experiment, but it could lead to further considerations, e.g. the choice of a higher AF (it is noted that the current AF range for mesocosms does not address the potential lack of representativeness) and/or a request for further testing to confirm that EPT are not among the most sensitive species. In this assessment, particular consideration should be paid to the mode of action of the active substance (e.g. mayflies are particularly sensitive to neonicotinoids, while other chemicals are known to affect other taxa more).

**Experimental design**

*Establishment time*: the pre-treatment period should be sufficient to allow the populations and communities to be well-established in the system before the first treatment. If this period is too limited, it can lead to low abundance of some (sensitive or vulnerable) populations which will make any effects more difficult to detect.

*Recolonisation*: insect recolonisation of treated mesocosms is sometimes due to control mesocosms acting as 'source' in the immediate vicinity (i.e. not necessarily representative of actual field conditions, where 'clean' bodies of water may not be present). This may result in an overestimation of the recovery potential. Hence, careful consideration of the life cycle of the taxa under investigation should be made, where possible, in order to assess the influence of this process.

*Emergence*: when emergence is the relevant endpoint, the cumulative emergence over time (when applicable) should also be provided, calculated, and graphically presented. This is particularly relevant when no suitable measurement of the aquatic population size is available, and therefore emergence is used as a proxy for the population size. Considerations about voltinism and synchronicity are also relevant for a proper assessment.

*Insect instar*: for insects, it is also important to check which instars are added at the beginning of the test; there are examples (e.g. studies with EPT or other insects) where the wrong instars were used (e.g. close to emergence), resulting in very minimal exposure and high uncertainty in the outcome of the test.

*Replicates*: three replicates per treatment is ideal but mesocosms studies are often performed with only two replicates. This results in limitations in the interpretation of results, as abundance is variable and detection of effects is thus limited.

*Number of samples*: at one sampling time, the number of samples taken within a replicate could be increased in order to decrease the variability. However, as sampling is in general destructive, one must pay attention not to deprive the mesocosm of too many organisms, thus inducing an 'artificial' stress to the model ecosystem.

*Sampling times*: early sampling frequency for short-life-span species and fast degrading substances can be critical because a fast recovery could hinder the detection of short but pronounced effects (thus misinterpreting a class 3A effect as a class 1 effect). As a general rule, sampling with a 7-day frequency or lower (e.g. sampling at 0, 7, 14 days after treatment) should be considered with care. It must be noted that the setting of the effect classes (and hence the NOEC) is influenced by the number of consecutive sampling dates when an effect is detected.

**Effect classes**

The terminology for effect classes currently included in the (EFSA PPR Panel, 2013) is based on the definitions by Brock et al. (2006) and De Jong et al. (2008) and modified to add the information about the minimum detectable difference (MDD).

Effect class 2 (slight effects) is defined as 'Effects concern short-term and quantitatively restricted responses usually observed at individual samplings only'.

MDD classes do not propose a quantification for 'slight effects', but they do set to 50 % the limit for MDD able to detect 'small effects' (MDD class IV).

Brock et al. (2015) suggested that a class 2 effect can be set if the MDD is < 70 % on the sampling after the effect, or < 90 % on the two samplings after the effect. The paper also added that class 2 effects can be set when, on the sampling after the effect, the percentage deviation from controls is less than 20 %.

It must be noted that the decision scheme in the (EFSA PPR Panel, 2013) for the setting of the NOEC on the basis of effect class 2 concentration does not specify an MDD trigger nor a proper percentage effect for the sampling times following the one indicating an effect. This indeed opens up possible interpretation on the criteria to be used for setting class 2 effect concentrations. This should be further clarified in the revision of the (EFSA PPR Panel, 2013).

**Consideration of indirect effects**

Community interactions (indirect effects; food chain effects) are to be appropriately considered when assessing effects of PPPs. For example, if the recovery option is selected for algae in a study with a herbicidal mode of action, the study should be critically evaluated for potential effects on higher trophic levels (e.g. zooplankton).

**Definition of population experiments as Tier 3 approach**

It is questionable why a population experiment performed with a single species (e.g. single species fish studies) would be submitted as a higher tier (Tier 3) approach rather than as an intermediate tier (e.g. Tier 2 approach, refined exposure test system, Tier 2C).

The 'population experiment' mentioned as a Tier 3 in the tiered approach, means that the focus is on a specific population within a community. Similarly, if a mesocosm experiment was performed, for example, with a very reduced number of species, the same concern would apply.

The definition of a micro-/ mesocosm is sometimes problematic. Irrespective of the definition, from a risk assessment perspective, any Tier 3 experiment (microcosm or mesocosm, indoor or outdoor) is a test system that includes an assemblage of species. In the case of microcosms, this could be of smaller size and/or shorter time duration than a mesocosm, but nonetheless it should include a representative community, where interaction between species is considered.

### 4.4. Representativeness of mesocosm studies when the risk assessment at lower tiers is triggered by a non-freshwater species

The current aquatic guidance (EFSA PPR Panel, 2013) was developed to perform risk assessments for freshwater environments, in accordance with the data requirements specified in EU Regulations 283/2014 and 284/2013. The same AGD, however, does not exclude the opportunity of using data from non-freshwater (marine or brackish) species in the risk assessment scheme. On the contrary, endpoints for these species are regularly used in the evaluations of active substances and PPPs.

Data from ecotoxicological tests on non-freshwater species can refer to species at all trophic levels (e.g. *Skeletonema costatum* for primary producers, *Americamysis bahia* for aquatic invertebrates and *Cyprinodon variegatus* for fish). It is not unusual that the lower tier risk assessment is driven by non-freshwater species. When the evaluation at these lower tiers highlights a potentially high risk, an option to refine the assessment is to conduct mesocosm studies on freshwater communities. Non-freshwater species are hardly represented in such mesocosms, and therefore it is questionable whether these studies are adequate to derive an endpoint able to cover the organisms represented at lower tiers by non-freshwater species.

Usually, the presence of other organisms considered taxonomically similar to the most sensitive non-freshwater species is taken into account to solve the issue. However, the concept of 'taxonomically similar' is open to many interpretations: the term 'taxon' indicates a group of organisms with similar characteristics that can be applied to all the hierarchical levels of biological classification.

The role of phylogeny was discussed at the meeting and some experts disagreed about the use of this approach. It was highlighted that phylogeny is very fluid and hence difficult to be relied upon.

The proposal of setting a 'fixed' taxonomic hierarchical limit is problematic, as for some groups it is possible to get a better picture (more sub-group represented) than for others. However, a minimum

level to be addressed was proposed on the basis of the comparison between *A. bahia* and the more closely related taxa that are often tested in mesocosms (Gammarids and Isopods). On this basis the minimum level to be matched should be the superorder. However, a general rule should be to consider which is the closest taxon that can reasonably be tested in a mesocosm, considering its autecology.

Overall, a stepwise procedure was proposed and agreed upon:

**Step 1**: check whether in the mesocosm the taxa closely related to *A. bahia* are included as the minimum representativeness requirement.

- If the mesocosm does not meet the minimum representativeness requirement, it cannot be considered to cover the risk for the most sensitive taxonomic group.
- If the mesocosm covers the minimum representativeness requirement, go to step 2.

**Step 2**: check that the 'representative surrogate taxa' (those taxonomically similar to the marine species driving the risk assessment at Tier 1) respond to the treatment, showing clear effects.

- If the 'representative surrogate taxa' respond to the treatment, the mesocosm is considered representative and can be used to address the risk assessment.
- If the 'representative surrogate taxa' do not respond to the treatment, go to step 3.

**Step 3**: perform further analysis and additional laboratory experiments might be requested with the 'representative surrogate taxa'. This would allow a better interpretation of the mesocosm by verifying whether the sensitivity of the 'representative surrogate taxa' is similar to that of the marine species untested in the mesocosm.

### 4.5. Use of refined exposure studies as Tier 2C

At the meeting, Germany presented an update on the central zone harmonisation meeting regarding the use of refined exposure studies. A position paper was also made available before the meeting. Nevertheless, it was pointed out that a complete agreement could not be reached at the central zone level regarding these kinds of experiment. Representatives from the northern zone reported that this kind of refinement is not considered acceptable for their zonal assessments. It was explained that this is mainly due to doubts that the FOCUS profiles can accurately reflect exposure in the field (particularly as they are currently based on limited time simulations). It was, however, noted that the same doubt should also apply to the use of mesocosms, for which exposure profiles are also compared to the FOCUS predictions. Other concerns were related to the uncertainties in the extrapolation of the results to the field, e.g. the uncertainties on the life stage of the tested species which are exposed in this kind of test. It was indeed highlighted that it is very difficult to have a match of the pulsed exposure with the most sensitive life stage, particularly when knowledge is lacking about which is the most sensitive stage.

It was also noted that the use of the Tier 2C refinement may be problematic for populations of short-lived species (e.g. algae, aquatic plants, daphnids). Indeed, some potential recovery may take place in these tests, while ERO is not an option at Tier 2, as recovery in the field would be influenced by the relationship with other species. For primary producers, it was suggested that an $EC_{10}$ be used instead of an $EC_{50}$, in order to reduce the possibility of an effect that it is 'absorbed' by a subsequent recovery (it should be noted that this approach is already included in the position paper presented by Germany). In addition, repeated measurements over time of the relevant endpoint(s) help to detect whether a possible recovery takes place. For daphnids and other short-lived invertebrates, testing at the individual level (i.e. not using populations) should exclude any concern about recovery at the population level, since only repair mechanisms at the level of the individual occur.

In the approach (still not agreed) initially suggested for the central zone, a prerequisite for carrying out refined exposure tests is to provide a risk assessment using endpoint(s) from experiments carried out under constant exposure and that includes mitigation measures. Everyone agreed that providing a lower tier risk assessment with mitigation measures is a reasonable approach for all kinds of refinement. However, it was also highlighted that this does not relate specifically to Tier 2C in any way. It was also agreed that showing a low risk with mitigation measures at lower tiers should not be considered as a reason to avoid an assessment of the available higher tier studies.

All experts agreed that the scheme for assessing Tier 2C should be reconsidered and possibly further developed in in the revision of the AGD.

### 4.6. Alternative test design in *Myriophyllum* studies

OECD Test Guidelines (TG) 238 and 239 (OECD, 2014a,b) describe the test designs to perform toxicity tests with the rooted aquatic dicotyledon *Myriophyllum spicatum* in the absence and presence of sediment, respectively. Both test guidelines require at least five tested concentrations (plus the control) for the determination of the $EC_X$. Test Guideline 238 requires 10 replicates for the control(s) and five replicates for the tested levels, with a single lateral branch for each replicate. Test Guideline 239 requires instead a minimum of six replicates for the control(s) and a minimum of four replicates for the tested levels; each replicate, represented by a test vessel, is composed of three shoots that can be managed in accordance with one of two test designs:

- Test Design A: one shoot per pot and three pots per vessel
- Test Design B: three shoots per pot and one pot per vessel.

Test Guideline 239 reports that 'Alternative test designs of one shoot per pot per test vessel are acceptable provided that replication is adjusted as required to achieve the required validity criteria'.

An alternative test design has been used in toxicity tests for (at least) three active substances: halauxifen-methyl (EFSA, 2014b), florpyrauxifen-benzyl (EFSA, 2018) and oxasulfuron (EFSA, 2017b). In each test, a single shoot was used for each replicate, but the number of replicates was increased to 10 for the control and to five for the tested levels. The studies with this modified test design were considered acceptable in two cases (halauxifen-methyl and florpyrauxifen-benzyl) but were rejected in the third since the number of individuals was considered too low.

The comparison of the two test designs (i.e. the one reported in the OECD test guidelines and the one with single shoots per replicate) (Gonsior and Schwalbach, 2014) gave very consistent results.

The use of single shoots in each replicate allows the use of 'real' replicates without interaction among individuals and to increase the statistical power of the test, particularly for the control, owing to a higher number of replicates. Given that the proposed alternative test design is in line with the OECD TG 238, the experts at the meeting agreed to consider it acceptable, as proposed by Italy.

### 4.7. How to express the endpoint for sediment-dwelling organisms when tested in the presence of sediment

During the Pesticide Peer Review Meeting 133 (EFSA, 2015) it was discussed how the endpoints for aquatic Tier 1 studies should be expressed. It was agreed that 'the toxicity endpoint for Tier 1 studies (i.e. mean measured, nominal or initial measured), should not depend on the study design, on the physical chemical or environmental fate parameters, on technical difficulties when testing, or on how the endpoint would be used in the first-tier risk assessment. The choice must depend on the actual exposure throughout the whole exposure period of that particular test. Where a suitable exposure throughout the whole period was not demonstrated, none of the endpoints should be used in first-tier risk assessments.' This discussion did not specifically cover the case of the toxicity tests on sediment-dwellers when tested in the presence of sediment.

The studies more frequently available for addressing the effects on sediment dwellers are performed on *Chironomus riparius* (OECD 2014a,b).

According to OECD TG 218 (sediment–water chironomid toxicity using spiked sediment), in order to assess the behaviour/partitioning of the tested chemical in the water–sediment system, the concentrations of the test substance should be measured in the sediment, in the pore water and in the overlying water. These analytical determinations indeed allow for the calculation of mass balance and to express the results based on measured concentrations. According to the same guideline, effect concentrations should be expressed and based on dry weight and preferably based on measured sediment concentrations at the beginning of the test (OECD, 2004a). Further recommendations on how to express the endpoint in the cases where the test item concentrations are not maintained (considering the whole system) or on how the mass balance results should be considered in this context are not included in the test guideline.

Similarly, according to OECD TG 219 (sediment–water chironomid toxicity using spiked water), samples of the overlying water, the pore water and the sediment must be analysed in order to assess the behaviour/partitioning of the tested chemical in the water–sediment system. The test guideline recommends that effect concentrations are expressed as concentrations in the overlying water, preferably calculated based on measured concentrations at the beginning of the test (OECD, 2004b). Further recommendations on how to express the endpoint in the cases where the test item concentrations are not maintained (considering the whole system) are not included in the test guideline. In addition, differently from OECD TG 218, there is no recommendation to calculate a mass balance in order to assess the behaviour of the test item in the system.

In the context of the peer review of the active substance risk assessment, the issue of how the concentrations should be expressed in the case of sediment-dweller toxicity testing was often raised. In particular, there have been instances in which it was questionable to express the endpoints as measured concentrations at the beginning of the test, i.e. in the cases where the concentrations were not maintained in the whole system.

EFSA recommended that the decision on how to express the endpoint for the sediment-dwellers is based on the assessment of the mass balance calculation in order to determine the repartition of the substance in the various compartments. In this view the submission of mass balance calculations as part of the dataset for the sediment-dwellers is highly recommended, particularly in the case of the substances that are difficult to test (concentrations poorly maintained in the test system). In the latter cases, it is also relevant that intermediate measurements in the various compartments are performed (see also Regulation (EU) No 283/2013, Section 8.2.5.3). When a mass balance is available, it is possible to consider the recommendations of the Pesticide Peer Review Meeting 133 (EFSA, 2015). It is additionally recommended that the key endpoints from the sediment-dweller studies are always presented in terms of mg substance/kg dry sediment and mg substance/L water. This would ensure that both exposure via water and sediment are covered for sediment-dwellers.

Where the concentrations in the test system are not maintained, the recommendations of the Pesticide Peer Review Meeting 133 (EFSA, 2015) should be considered, i.e. express the endpoint as the mean measured concentration using mg substance/kg dry sediment and/or mg substance/L water, accordingly, if significant levels are detected in the sediment or in the water or in both. The calculations should be based on geometric mean concentrations. It is proposed to further discuss whether, in such cases, the use of these studies in a Tier 2C approach, similar to the proposal in the EFSA aquatic guidance document (EFSA PPR Panel, 2013) for the refined exposure studies, would be suitable. This means that it should be demonstrated that the exposure in the study simulates a realistic worst-case exposure relative to the predicted exposure. In this view, a comparison between the exposure in the test system and the expected exposure (FOCUS profiles) should be performed. In order to follow this approach, intermediate analytical measurements should be performed in the course of the study.

It is acknowledged that issues similar to those for the sediment-dwellers could also occur for toxicity tests with the rooted macrophyte *Myriophyllum spicatum* (OECD TG 239; OECD, 2014b). In those cases it is suggested that the same approach as above is applied. It is noted that OECD TG 239 already highlights that 'if there is evidence that the concentration has declined (i.e. is not maintained within 20 % of the nominal or measured initial concentration in the treated compartment) throughout the test, then analysis of the results should be based on the geometric mean concentration during exposure or models describing the decline of the concentration of the test chemical in the treated compartment'.

Overall, the experts agreed with the proposal to use the mass balance for checking whether the concentrations were adequately maintained. Practical examples of the needed calculations are included in Appendices G and J.

## 4.8. Other issues on aquatic organisms

**Minimum detectable difference**

The MDD, presented in the (EFSA PPR Panel, 2013) and the paper by Brock et al. (2015), is considered to be a valid tool to help with the evaluation of the biological results to assess the statistical power – or the absence of power – of a study to detect treatment-related direct effects. It

should preferably be reported on non-aggregated data for the relevant taxon and time points. An issue linked to the unclear beta-error associated with the MDD in the available documents (i.e. (EFSA PPR Panel, 2013) and Brock et al., 2015) was raised by Germany.

It was concluded that the use of the MDD is supported and that further considerations and clarifications will be addressed in the revision of the (EFSA PPR Panel, 2013).

**Primary producers**

Some concerns have been raised regarding the level of protection reached when using the $ErC_{50}$ values for risk assessment for algae and aquatic plants since the implementation of the AGD. The point was proposed for discussion by Germany, who presented a meta-analysis of Tier 1 and higher tier data. It was shown that Tier 1 endpoints expressed in terms of growth rate (i.e. $ErC_{50}$ values) for algae and *Lemna* are respectively 6.9- and 3.5-fold higher than the $Eb/yC_{50}$ values. Furthermore, comparison of Tier 1 data with endpoints from mesocosm studies indicated that the Tier 1 RAC calculated using $ErC_{50}$ values is only protective in 42% of cases; while the same comparison based on $EbR_{50}$ indicated a sufficient level of protection in 75% of the cases.

The experts acknowledged this concern. However, considering the available scientific knowledge, it was suggested that EFSA further consider this issue in the context of the revision of the (EFSA PPR Panel, 2013) by taking into consideration all the available scientific knowledge on this aspect (van Wijngaarden and Arts, 2018).

# 5. Non-target arthropods and soil organisms

Several issues related to the risk assessment of NTAs which emerged during the peer review of the active substance, were proposed for discussion both by EFSA and Member State experts. The main points are listed below:

- Aged residue trials and recovery
- Minimum time considered acceptable for an in-field recolonisation
- Vegetation distribution factor (VDF)
- Risk assessment for NTA when contact exposure is not relevant
- Whether the aspects of the de Jong Guidance et al. (2010) should be used and whether extrapolation of field studies between crops is appropriate.

As regards the risk assessment for earthworms, the point raised was on the use of de Jong Guidance et al. (2010) for evaluating and summarising field studies.

The background and the meeting conclusion on the above points are reported below.

## 5.1. Aged residue trials and recovery

This point for discussion was proposed by countries in the northern zone. The evaluation of the environmental risk of pesticides to NTAs is conducted according to a tiered approach consisting of:

Tier 1 tests: Exposure of two sensitive species of arthropods (*Aphidius rhopalosiphi* and *Typhlodromus pyri* ) to the formulated product applied to glass plates followed by the hazard quotient approach.

Tier 2 and 3 (higher tier) studies: Extended laboratory tests with fresh dried residues (Tier 2) or 'aged' residues (Tier 3): these are tests with natural substrate (leaves or whole plants) intended to measure lethal and sublethal effects. Test organisms are applied on the leaves or plants at either day 0 after treatment (Tier 2) or several days after treatment (Tier 3). In Tier 3 studies, the leaves (2D tests) or whole plants (3D tests) are sprayed and left under normal weather conditions to age for a variable duration. Tier 3 studies aim to provide information on the time scale needed for the potential for recolonisation of treated areas.

For extended laboratory studies, a protocol is available for *A. rhopalosiphi* (Mead-Briggs et al. 2010 – only on freshly dried residues) involving the use of barley seedlings (3D test) since *A. rhopalosiphi* is mainly a parasite of grass aphids. However, for *T. pyri* and other test organisms no protocols are available for extended laboratory tests. Only Tier 1 (glass plates) protocols are available in the report

from the SETAC/ESCORT 2 Workshop from Candolfi et al. (2001). As a result, a variety of test systems (bean or vine leaves, maize plants, etc.) are applied in extended laboratory studies.

The relevance of the test system for the specific good agricultural practice (GAP) under assessment is often not straightforward to interpret.

The following two issues were highlighted for discussion:

1. Whether the substrate used for the aged residue study matters and whether the substrate should be relevant to the GAP.

2. The allowable ageing duration that can be considered to show potential for recovery (discussed under Section 5.2 below).

A concern was raised regarding the substrate used in the higher tier studies. Some experts believed that this extrapolation is not a source of major concern. Other experts expressed concern about the lack of standardisation of the substrate. It was pointed out by some experts that differences in substrate had never raised a concern in the past and there seems to be no clear evidence that the different substrates are yielding different test results. The type of plant which is used in the higher tier studies with NTAs depends also on the preferences of the NTA species for certain plants and the practicability of testing. The experts did not reach an agreement regarding the issue of test substrate. Therefore, it was suggested that this is considered in the context of the revision of the guidance document for NTAs.

## 5.2. Minimum time considered acceptable for an in-field recolonisation

When the Tier 1 risk assessment for NTAs indicates a high risk, a higher tier risk assessment is needed to evaluate the potential for recovery of NTA populations, i.e. the return of an ecological entity to a defined reference state after a disturbance. This can be done using aged residue studies, combining information from extended laboratory studies (with fresh residues) and information from the degradation of the product, or with semi-field or field studies. However, while semi-field or field studies account for actual recovery, aged residues and a combination of information can only assess the possibility of a potential recovery.

In the current risk assessment scheme, the recovery option is considered only for the in-field evaluation, since it is supposed that the off-field acts as a reservoir for the in-field recolonisation. In-field recovery can indeed result from processes internal to the field (growth and reproduction of surviving individuals) or from immigration of individuals from the off-field. While the internal recovery can be de facto completely reset by the application of a pesticide, the option for recolonisation from the off-field is always guaranteed (i.e. low off-field risk). The rates of these two processes depend on species characteristics, such as the number of generations per year or life cycle strategies on one hand, and dispersal capacity or territorial behaviour on the other hand. Consequently, the time needed for the complete recovery of a population can vary widely among species.

The current risk assessment for NTAs, based on the Guidance Document on Terrestrial Ecotoxicology (European Commission, 2002) and, in turn, on the outcome of the SETAC/ESCORT 2 workshop, suggests a period of one year as the limit for the recovery to be considered acceptable, without distinction between potential and actual recovery. While the actual recovery represents an evidenced recovery, the potential recovery indicates that detrimental effects caused by a PPP after a certain period fall below the limit of 50%, without any indication that a recovery in fact takes place. Consequently, the internal recovery of species with low reproductive capacity and/or the recolonisation of species with low mobility could take a longer time. Moreover, during the peer review process, recovery times shorter than one year were considered from time to time.

The following issues were proposed for discussion at the meeting:

- The acceptable time for recovery/recolonisation considering the potential and the actual recovery time from aged residues, semi-field and field studies (i.e. should different recovery times be accepted for data from different studies?).
- Agronomical practices can influence the time needed for a complete recovery, prolonging it. For example, foliar insects in an arable crop severely affected at the beginning of the growing

season probably take more than 12 months to recover, since plants are harvested at the end of the growing season and during winter the biological activities are stopped.

- To demonstrate the potential for recovery, different times could be considered acceptable on the basis of data from extended laboratory studies, i.e. if effects on mortality and reproduction are >50% but <100% in an extended laboratory study, a different recovery time could be considered in comparison to an extended laboratory study with 100% mortality, since in the first case a subsample of the initial population can constitute a core for an internal recovery.

Due to time constraints, the discussion in the meeting focused on the first point above on the acceptable time for recovery/recolonisation and the actual recovery time from aged residues, semi-field and field studies; see Sections 5.2.1 and 5.2.2, below.

### 5.2.1. Aged residues

Some experts suggested that the allowable ageing duration that can be considered to show potential for recovery should be lower than the generation time or life span of the tested species and a duration of ageing of < 1 week was proposed for *Typhlodromus pyri* and ≈ 1 week for *Aphidius rhopalosiphi*. It was noted that this is a very short period of ageing in comparison to the requirement in the NTA assessment to show the potential for recovery and recolonisation within one year and other, longer time periods for ageing of residues were proposed by some experts. However, it was not possible to agree on a proposal for the duration of ageing since a sound scientific basis for such a proposal was missing. Further work would need to be done on the bio-ecological traits of the species being tested in order to identify the time allowed for ageing. It was noted that the test species might not be a good 'ecological' representative of the community present in the crop to be treated. Furthermore, the mode of action of a PPP and the GAP (time of application and number of applications) needs to be taken into consideration when deciding on the maximum time of ageing. Hence, it was agreed that setting a duration limit for aged residue trials in the context of the meeting may lead to an arbitrary decision and that this issue should be addressed in future guidance documents. The experts also raised the additional concern that the use of aged residue studies, in general, may not be sufficient to demonstrate recovery/recolonisation as recolonisation is highly dependent on the landscape configuration.

### 5.2.2. Semi-field and field studies

The discussion on acceptable recovery times for semi-field and field studies was not discussed in depth. Similar issues to those which were raised for the aged residue trials are also applicable for semi-field and field studies. Concerns were raised by the experts that the acceptability criterion of the potential for recovery/recolonisation within one year is not sufficiently protective.

## 5.3. Vegetation distribution factor

Currently, a VDF of 10 is used in the risk assessment for NTAs as proposed by the report of the SETAC/ESCORT 2 Workshop (Candolfi at al., 2001) based on 'leaf area indices' and 'plant interception'. According to (European Commission, 2002), 'this figure is considered unreliable, therefore more appropriate data should be used as soon as they become available'.

Several reviews of the VDF value and attempts to derive an appropriate default figure for the VDF are available and all these evaluations were presented in Appendix E of the EFSA scientific opinion on NTAs (EFSA PPR Panel, 2015). These reviews indicate that a VDF of 10 is not appropriate.

It was proposed by some Member States of the central zone to specifically discuss the following options:

- Not to use a VDF at Tier 1
- To use a VDF of 3–5 at higher tiers.

Overall, the majority of the experts agreed on the recommendation of using a VDF of 5 for all the tiers of the assessment. It was highlighted that this recommendation should be considered as an interim solution until the revision of the current risk assessment scheme. Such an interim solution should be

reflected in the (European Commission, 2002) document and its implementation should be further considered.

## 5.4. Risk assessment for non-target arthropods when oral exposure is relevant

Oral exposure via herbivorous NTAs is not covered by the current standard risk assessment methodology. The EFSA NTA Opinion (EFSA PPR Panel, 2015) recommends adding a herbivorous species at Tier 1 of the risk assessment in order to also cover oral uptake and herbivorous arthropods. In particular, Lepidoptera are suggested as a test organism and *Pieris brassicae* was considered by the EFSA PPR Panel (2015) as a suitable organism for testing but no standardised test guidelines are available.

It was not considered appropriate to suggest a new test species within the context of this technical report. However, it was considered important to highlight the potential risk from oral exposure when it is relevant. In cases where the active substance is targeted toward sucking or biting herbivorous insects it is commonly understood that the current risk assessment methodology does not cover exposure to herbivorous NTAs. Therefore, it was agreed that a concern should be noted in the EFSA conclusion on a case-by-case basis (e.g. for substances which are targeted against sucking and biting herbivorous insects).

## 5.5. Use of de Jong et al. (2010) guidance for non-target arthropod field studies

Currently there is no agreed guidance at the EU level for the evaluation of NTA field studies. This may lead to differing evaluations at EU level and frequently to discussion points in experts' meetings. Harmonisation of the evaluation of field studies would therefore be beneficial. A possible option would be to use aspects of the de Jong guidance (de Jong et al., 2010) which is also suggested by the EFSA NTA Opinion (EFSA PPR Panel, 2015) as a guideline for summarising and evaluating NTA field studies until further guidance is developed.

It was proposed by EFSA to start using aspects of the de Jong et al. (2010) guidance for EU-level assessments in order to have a more harmonised assessment of higher tier NTA studies. The experts acknowledged that using the guidance from de Jong et al. (2010) has advantages and that some aspects of the guidance should be used for EU-level assessments until further guidance is available for evaluating NTA field studies. The elements agreed have been in included in a template in Appendix H. It is recommended that this template is followed when reporting the studies in the RARs/DARs.

In using the guidance, the experts agreed that the level of aggregation/detail as proposed in Table 2 of de Jong et al. (2010) is useful for summarising the results. Consequently, it should be included in the study summary presented in the RAR/DAR. All experts agreed that the taxa listed in Table 4 of de Jong et al. (2010) should be used as a reference for the reliability assessment. Footnotes to Table 4 were missing from the guidance and EFSA contacted the author who made them available. For ease of reference, the footnotes to Table 4 are summarised at the end of Appendix H of the current report. It was agreed that further information and argumentation should be presented when specific taxa are missing in the field study. The experts also agreed that studies should include a toxic reference item or to apply rates of the test item high enough to cause clear effects. If a suitable toxic reference was not available, unless effects were clearly seen with the test item, the study should be classified as 'unreliable'. The experts agreed that presenting the results in terms of effect classes as suggested by de Jong et al. (2010) are recommended but should not be considered mandatory.

## 5.6. Use of de Jong et al. (2006) guidance for earthworm field studies

Earthworm field tests are carried out according to ISO 11268-3 (2014). In 2006, guidance on how to summarise those studies was published by de Jong et al. (2006).

The guidance gives recommendations on a number of items which should be considered when assessing the reliability of an earthworm field study. EFSA proposed to adopt the approach described in this document for summarising and evaluating the earthworm studies in the RARs/DARs. Overall,

the experts at the meeting agreed that the recommendations from de Jong et al. (2006) could be very useful, but some modifications were proposed. The elements agreed have been included in a template that is provided in Appendix I. It is recommended that this template is followed when reporting the studies in the RARs/DARs.

### 5.7. Use of the minimum detectable difference for interpreting field studies on non-target arthropods and earthworms

The MDD is considered by the experts as a valid tool for evaluating the biological results. Although it could give some information for the assessment of higher tier studies, overall it was considered premature to recommend calculation of the MDD for higher tier studies performed with NTAs and soil organisms, because criteria to help interpret these MDD values are currently lacking (e.g. classes of MDD, minimum number of taxa with an acceptable MDD).

## 6. Non-target terrestrial plants

This issue was proposed and presented by a representative from the central zone. In addition to seedling emergence, OECD T 208 (OECD 2006a) and vegetation and vigour, OECD TG 227 (OECD 2006b), other variables, such as visual phytotoxicity, and sometimes shoot length, are evaluated according to these respective guidelines. $ER_X$ values for visual observations (also referred to as 'visible detrimental effects' or 'visual injury', such as chlorosis, necrosis, wilting, leaf and stem deformation) could be determined, where a dose–response relationship is available, but this is not often the case. The experts at the meeting discussed the relevance of using this endpoint in the Tier 1 risk assessment. The experts considered that effects on growth may also cover the phytotoxicity endpoint, which may be subjective being based on visual assessment. However, it was noted that the EFSA PPR Panel (2014) reported that for a significant number of cases this endpoint was reported as being lower than the others. Therefore, considering that the endpoint is part of the test guidelines and that the data requirements do not specify the parameters to define the endpoint for risk assessment, the experts concluded that the $EC_X$ based on phytotoxicity should be reported in the study summary and in the list of endpoints. Where the derived endpoint is the lowest of those available, it should be considered for the Tier 1 risk assessment. Such an interim solution should be reflected in the (European Commission, 2002) document and its implementation should be further considered.

## 7. Overall conclusions and recommendations

General and specific issues related to risk assessment for birds and mammals, aquatic organisms, NTAs, soil organisms and non-target terrestrial plants were identified and discussed in a general ecotoxicology meeting, the Pesticide Peer Review Meeting 185, which took place from 9 to 12 October 2018.

Recommendations on these topics were compiled based on the discussion and conclusions achieved at the meeting and they are summarised below in Table 1. Many of these recommendations can be applied during the EFSA peer review of the active substances and they are expected to provide additional clarifications to applicants and rapporteur Member States regarding the scientific interpretation of the relevant issues when preparing the dossiers and the assessment reports. Furthermore, it is expected that some recommendations will be taken into consideration during the revision of the relevant guidance documents in the area of ecotoxicology.

**Table 1**: Overview of the recommendations compiled based on the discussion and conclusions achieved at the general ecotoxicology meeting, Pesticide Peer Review Meeting 185

| | |
|---|---|
| **Extrapolation of studies between different agroclimatic conditions** | For aquatic mesocosm studies, the simplest approach for risk assessment, in order to cover the extrapolation between different climatic areas, is to perform the risk assessment using the ETO RAC approach. The use of the ERO RAC approach across agroclimatic zones needs to be considered on a case-by-case basis.<br><br>For birds and mammals, any refinements of the risk based on |

| | |
|---|---|
| | identification of specific focal species and definition of related ecological data should use data which are representative for the area of use of the active substance; this need is already reflected in EFSA (2009). |
| **How to consider studies when the analytical methods are not validated** | For studies where the analytical method used for the chemical analysis is assessed to be not fit for purpose, a case-by-case evaluation of the reliability of the endpoints derived from the study should be conducted. This should consider all the available information, including the toxicological profile of the substance and the margin of safety of the risk assessment, before rejecting the study. Additional vertebrate testing should be avoided where possible. If an endpoint from a study with an analytical method assessed to be not fit for purpose is deemed sufficient for risk assessment, this should be clearly indicated in the list of endpoints . |
| **Risk assessment for PPPs within the evaluation of the active substance** | To decide whether an endpoint from a study performed with a PPP indicates greater toxicity relative to the active substance, the experts at the meeting recommended using a factor of 3, i.e. when the endpoints are both expressed in terms of the active substance, if the endpoints are within a factor of 3, then this can be considered as inter-study variability.<br><br>In cases where the PPP indicates greater toxicity relative to the active substance (i.e. > a factor of 3), in accordance with the data requirements, the lower endpoint should be used for risk assessment. Alternatively, separate risk assessments for both the active substance and the PPP could be performed.<br><br>For aquatic toxicity studies with PPPs containing multiple active substances, unless it is clear which substance drives the toxicity, it is recommended that analytical measurements should be performed for all active substances in the PPP. In the case that the concentration of one of the active substances was not maintained, then the study should not be used for Tier 1 risk assessment, but it could potentially be used in a Tier 2 risk assessment. |
| **Use of residue data to support ecotoxicological assessments** | There was a general consensus at the meeting that there is a need to establish an efficient and consistent communication between the residue and the ecotoxicological sections.<br><br>As agreed, a questionnaire was developed to consistently collect and report the relevant information (see Appendix B). It is suggested that the information collected is reported via the questionnaire as an Appendix to section B.9 of the DAR/RAR. |
| **Equivalence of batches** | In cases where the test item used in studies has less than 90% purity and the endpoints are expressed in terms of nominal concentrations, the toxicity endpoints should be corrected for the purity of the test material. To be in line with the GAP and subsequent exposure assessments, this correction should be made to present the endpoints in terms of pure active substance.<br><br>It was agreed that an overview of the batches used in all the available ecotoxicological studies should be presented in Vol. 3 B.9 in line with the Commission guidance (European Commission, 2012): a Tier 1 assessment should be presented for all the batches used in the ecotoxicological studies while a Tier 2 assessment should only be performed for those batches used in |

| | |
|---|---|
| | key studies (i.e. studies used for risk assessment).<br><br>It was agreed that, in general, the non-compliance of the batches used in the ecotoxicity studies to the technical specification is not of concern to warrant the identification of a critical area of concern for ecotoxicology in the EFSA conclusions. Consequently, in these cases only a data gap will be identified. In cases where the available information indicates a potential concern (e.g. impurity considerably more toxic than the active substance), then a critical area of concern in the EFSA conclusion may be needed. |
| **Use of lower limit ($EC_{10}$, $HC_5$) as endpoint in the risk assessment** | The experts at the meeting concluded that an update of the guidance given in Appendix F of the technical report (EFSA, 2015) was needed. Such an update is included in Appendix E of this report, which provides a synthesis of the whole process. As agreed, the recommendations reported in Appendix E for appraising, reporting and using $EC_X$ values and related confidence intervals in the risk assessment. |
| **Risk assessment for rice paddies** | Italy presented a summary of the draft guidance document that Member States from the southern zone are currently developing. It was noted that, owing to their unique characteristics, specific considerations for risk assessments for rice paddies are needed. Consequently, the experts expressed appreciation for the ongoing development. |
| **Risk assessment for bananas** | It was agreed that, in the absence of a specific crop category for bananas, the orchard crop group can be used as a surrogate Tier 1 scenario for bird and mammal risk assessment according to EFSA (2009). In cases where a higher tier risk assessment is performed using ecological information, the applicant should provide appropriate data to identify focal species relevant for the crop and area of envisaged use. |
| **Birds and mammals** | **Trials for residue decline**<br>Kinetic assessment: General principles for the kinetic assessment were agreed.<br><br>Extrapolation: Rules for extrapolation within and between defined item groups were agreed.<br><br>Plant material: It was agreed that in order to refine the default value for residue decline, residue trials should be performed at at least four sites per item and regulatory zone. However, it was also agreed that in some cases there may be a possibility to extrapolate between areas (e.g. northern France).<br><br>Invertebrates: no agreement regarding the minimum number of trials or sites was reached and this should be resolved by the ongoing working group for the revision of the EFSA Guidance (2009).<br><br>**21-day PT**<br>The experts agreed that, owing to the reasons discussed in Section 3.2, the methodology proposed in Ludwigs et al. (2017) for deriving 21-day PT values should not be used in risk assessment. It was recommended that the working group for the revision of the EFSA Guidance (2009) should reflect on this methodology. |
| **Aquatic organisms** | **Use of geometric mean and weight of evidence for acute** |

| | |
|---|---|
| | **data**<br><br>It was agreed that, in cases where the $RAC_{geomean}$ is greater than the lowest endpoint, the lowest endpoint should be used to calculate the $RAC_{lowest}$. The minimum modified AF for deriving the $RAC_{lowest}$ should be 20 for invertebrates and 30 for fish.<br><br>The experts suggested that the approach should be further considered with the revision of the EFSA PPR Panel (2013).<br><br>**Use of geometric mean and weight of evidence for chronic data**<br><br>There was no agreement for using a geometric mean for chronic data. This should be further considered together with the entire approach when the aquatic guidance (EFSA PPR Panel,2013) is revised.<br><br>**General recommendations on mesocosm experiments**<br><br>It was agreed that the absence or low abundance of vulnerable groups, i.e. EPT, should not necessarily result in the invalidation of the experiment. However, their absence should trigger the need for further considerations, e.g. the selection of a higher AF and/or request for further testing to confirm that EPT are not among the most sensitive species. In such assessment, particular consideration should be paid to the mode of action of the active substance.<br><br>Several recommendations for the experimental design, consideration of indirect effects and definition of Tier 3 experiments were discussed and agreed.<br><br>**Representativeness of mesocosm studies when the risk assessment at lower tiers is triggered by a non-freshwater species**<br><br>A stepwise approach was discussed and agreed (see Section 4.4).<br><br>**Use of refined exposure studies as Tier 2C**<br><br>It was agreed that the scheme for assessing Tier 2C should be reconsidered and possibly further developed in the revision of the EFSA PPR Panel (2013) AGD.<br><br>**Alternative test design in *Myriophyllum* studies**<br><br>It was agreed that *Myriophyllum* studies performed to OECD TG 239 (OECD, 20014b) but with an alternative test design (i.e. one shoot per pot per test vessel) should be considered acceptable.<br><br>**How to express the endpoint for sediment-dwelling organisms when tested in the presence of sediment**<br><br>It was agreed that endpoints for sediment-dwelling organisms, when tested in the presence of sediment, should be determined using a mass balance calculation. In this view the submission of mass balance calculations as part of the dataset for the sediment-dwellers is highly recommended, particularly in the case of the substances which are difficult to test (concentrations |

| | |
|---|---|
| | poorly maintained in the test system).<br><br>**Minimum detectable difference**<br><br>It was concluded that the use of the MDD is supported and that further considerations and clarifications will be addressed in the revision of the EFSA PPR Panel Guidance (2013).<br><br>**Primary producers**<br><br>The experts acknowledged the issue related to the use of $ErC_{50}$ versus the level of protection, but it was agreed to further consider it in the context of the revision of the EFSA PPR Panel Guidance (2013). |
| **Non-target arthropods** | **Use of de Jong et al. (2010) guidance**<br><br>The experts at the meeting acknowledged that using the guidance by de Jong et al. (2010) is useful and that some aspects of the guidance should be used for EU-level assessments until further guidance for the evaluation of NTA field studies is available (see Appendix H).<br><br>**Aged residue trials and recovery**<br><br>It was agreed that until further guidance is developed, the substrate used in the aged residue studies does not need to be relevant for the crop under assessment.<br><br>The experts agreed that the use of aged residue studies and their ability to demonstrate recovery should be further considered in the context of guidance document development. This consideration would need to include the appropriate length of the ageing period.<br><br>**Minimum time considered acceptable for an in-field recolonisation**<br><br>The experts at the meeting expressed a concern with the currently defined allowable period for recovery/recolonisation (one year). It was agreed that this would also need to be addressed in future guidance documents.<br><br>**Vegetation distribution factor**<br><br>The experts agreed that the VDF value should be changed as better data are now available. It was recommended that a VDF value of 5 is applied for all the tiers of the assessment as an interim solution. Such an interim solution should be reflected in the (European Commission, 2002) document and its implementation should be further considered.<br><br>**Risk assessment for non-target arthropods when oral exposure is relevant**<br><br>It was agreed that, until guidance is developed and adopted, data for herbivorous species should not be requested. In cases where a concern is raised (e.g. based on the mode of action of the active substance), then this should be highlighted in the risk assessment and acknowledged in the EFSA conclusion. |
| **Soil organisms** | **Use of de Jong et al. (2006) guidance**<br><br>It was agreed to follow the guidance from de Jong et al. (2006) with some revision for reporting the studies in the RARs/DARs |

| | |
|---|---|
| | (see Appendix I).<br>**Minimum detectable difference**<br>It was overall considered premature to recommend calculating the MDD for higher tier studies with soil organisms (and NTA), as criteria to help interpret these MDD values are currently lacking (e.g. classes of MDD, minimum number of taxa with an acceptable MDD). |
| **Non-target terrestrial plants** | **Endpoint based on phytotoxicity**<br>It was agreed that an endpoint based on phytotoxic effects should be reported in the study summary and in the list of endpoints. Moreover, such an endpoint should also be used in the risk assessment where relevant. Such an interim solution should be reflected in the (European Commission, 2002) document and its implementation should be further considered. |

## References


Andrade TO, Bergtold M and Kabouw P, 2017. Minimum significant differences (MSD) in earthworm field studies evaluating potential effects of plant protection products. Journal of Soils and Sediments, 17, 1706-1714.

Azimonti G, de Boer W, Galimberti F, Marchetto F, van der Voe H, 2015. Comparison of NOEC values to EC10/EC20 values, including confidence intervals, in aquatic and terrestrial ecotoxicological risk assessment. EFSA supporting publication 2015:EN-906. 274 pp.

Bayona Y, Brulle F and Boivin A, 2018. Regulatory risk assessment of soil organisms in field: dealing with earthworm community. Proceedings of the SETAC Europe Annual meeting, Rome, Italy.

Brock TC, Arts GHP, Maltby L and Van den Brink PJ, 2006. Aquatic Risks of Pesticides, Ecological Protection Goals and Common Aims in European Union Legislation. Integrated Environmental Assessment and Management, 2, E20–E46.

Brock TC, Hammers-Wirtz M, Hommen U, Preuss TG, Ratte HT, Roessink I, Strauss T and Van den Brink PJ, 2015. The minimum detectable difference (MDD) and the interpretation of treatment related effects of pesticides in experimental ecosystems. Environmental Science and Pollution Research, 22, 1160–1174.

Candolfi M, Bigler F, Campbell P, Heimbach U, Schmuck R, Angeli G, Bakker F, Brown K, Carli G., Dinter A., Forti D, Forster R, Gathmann A, Hassan S, Mead-Briggs M, Melandri M, Neumann P, Pasqualini E, Powell W, Reboulet J-N, Romijn K, Sechser B, Thieme Th, Ufer A, Vergnet Ch and Vogt H, 2000a. Principles for regulatory testing and interpretation of semi-field and field studies with non-target arthropods. Journal of Pest Science, 73, 141-147.

Candolfi MP, Blümel S, Forster R, Bakker FM, Grimm C, Hassan SA, Heimbach U, Mead-Briggs MA, Reber B, Schmuck R and Vogt H, 2000b. Guidelines to evaluate side-effects of plant protection products to non-target arthropods. IOBC, BART and EPPO Joint Initiative, Gent, Belgium: IOBC/WPRS. 158 pp.

Candolfi MP, Barrett KL, Campbell P, Forster R, Grandy N, Huet M-C, Lewis G, Oomen P A, Schmuck R and Vogt H, 2001. Guidance document on regulatory testing and risk assessment procedures for plant protection products with nontarget arthropods. Report of the SETAC/ESCORT 2 Workshop, Wageningen, The Netherlands, SETAC-Europe, Brussels, Belgium.

Čoja T, Zehetner K, Bruckner A, Watzinger A and Meyer E, 2008. Efficacy and side effects of five sampling methods for soil earthworms (Annelida, Lumbricidae). Ecotoxicology and Environmental Safety, 71, 552–565.

de Jong FMW, van Beelen P, Smit C and Montforts M, 2006. Guidance for summarising earthworm field studies. RIVM report 601506006/2006. Bilthoven, The Netherlands. 47 pp.

de Jong FMW, Brock TCM, Foekema EM and Leeuwangh P, 2008. Guidance for summarizing and evaluating aquatic micro- and mesocosm studies. RIVM Report 601506009, Bilthoven, the Netherlands, 59 pp.

de Jong FMW, Bakker FM, Brown K, Jilesen CJTJ, Posthuma-Doodeman CJAM, Smit CE, van der Steen JJM and van Eekelen GMA, 2010. A guidance document of the Dutch Platform for the Assessment of Higher Tier Studies. Guidance for summarising and evaluating field studies with non-target arthropods. National Institute for Public Health and the Environment (RIVM). RIVM report number 601712006/2010.

Dinter A, Oberwalder C, Kabouw P, Coulson M, Ernst G, Leicher T, Miles M, Weyman G and Klein O, 2013. Occurrence and distribution of earthworms in agricultural landscapes across Europe with regard to testing for responses to plant protection products. Journal of Soils and Sediments, 13, 278–293.

EFSA (European Food Safety Authority), 2009. Guidance on Risk Assessment for Birds and Mammals on request from EFSA. EFSA Journal 2009;7(12):1438. doi:10.2903/j.efsa.2009.1438

EFSA (European Food Safety Authority), 2013. EFSA Guidance Document on the risk assessment of plant protection products on bees (Apis mellifera, Bombus spp. and solitary bees). EFSA Journal 2013;11(7):3295, 268 pp. doi:10.2903/j.efsa.2013.3295

EFSA (European Food Safety Authority), 2014a. EFSA Guidance Document on clustering and ranking of emissions of active substances of plant protection products and transformation products of these active substances from protected crops (greenhouses and crops grown under cover) to relevant environmental compartments. EFSA Journal 2014;12(3):3615, 48 pp. doi:10.2903/j.efsa.2014.3615

EFSA (European Food Safety Authority), 2014b. Conclusion on the peer review of the pesticide risk assessment of the active substance halauxifen-methyl (XDE-729 methyl). EFSA Journal 2014;12(12):3913, 93 pp. doi:10.2903/j.efsa.2014.3913

EFSA (European Food Safety Authority), 2015. Technical report on the outcome of the pesticides peer review meeting on general recurring issues in ecotoxicology. EFSA supporting publication 2015:EN-924. 62 pp.

EFSA (European Food Safety Authority), 2016. Technical report on the outcome of the pesticides peer review meeting on general recurring issues in mammalian toxicology. EFSA supporting publication 2016:EN-1074. 24 pp.

EFSA (European Food Safety Authority), 2017a. Technical report on the outcome of the pesticides peer review meeting on general recurring issues in physical and chemical properties and analytical methods. EFSA supporting publication 2017:EN-1221. 19 pp. doi:10.2903/sp.efsa.2017.EN-1221

EFSA (European Food Safety Authority), 2017b. Conclusion on the peer review of the pesticide risk assessment of the active substance oxasulfuron. EFSA Journal 2017;15(3):4722, 25 pp. doi:10.2903/j.efsa.2017.4722

EFSA (European Food Safety Authority), 2017c. EFSA Guidance Document for predicting environmental concentrations of active substances of plant protection products and transformation products of these active substances in soil. EFSA Journal 2017;15(10):4982. doi.org/10.2903/j.efsa.2017.4982

EFSA (European Food Safety Authority), 2018. Conclusion on the peer review of the pesticide risk assessment of the active substance florpyrauxifen (variant assessed florpyrauxifen-benzyl). EFSA Journal 2018;16(8):5378, 21 pp. doi:10.2903/j.efsa.2018.5378

EFSA PPR Panel (EFSA Panel on Plant Protection Products and their Residues), 2013. Guidance on tiered risk assessment for plant protection products for aquatic organisms in edge-of-field surface waters. EFSA Journal 2013;11(7):3290, 268 pp. doi:10.2903/j.efsa.2013.3290

EFSA PPR Panel (EFSA Panel on Plant Protection Products and their Residues), 2014. Scientific Opinion addressing the state of the science on risk assessment of plant protection products for non-target terrestrial plants. EFSA Journal 2014;12(7):3800, 163 pp. doi:10.2903/j.efsa.2014.380

EFSA PPR Panel (EFSA Panel on Plant Protection Products and their Residues), 2015. Scientific Opinion addressing the state of the science on risk assessment of plant protection products for non-target arthropods. EFSA Journal 2015;13(2):3996, 212 pp. doi:10.2903/j.efsa.2015.3996

EFSA Scientific Committee, 2017. Guidance on the assessment of the biological relevance of data in scientific assessments. EFSA Journal 2017;15(8):4970, 73 pp. doi:10. 2903/j.efsa.2017.4970

Eisenhauer N, Straube D and Scheu S, 2008. Efficiency of two Widespread Non-Destructive Extraction Methods under Dry Soil Conditions for Different Ecological Earthworm Groups. European Journal of Soil Biology, 44, 141-145.

European Commission, 1997. Appendix A. Metabolism and distribution in plants. 7028/IV/95-rev.3.

European Commission, 2000. Residues: Guidance for generating and reporting methods of analysis in support of pre-registration data requirements for Annex II (part A, Section 4) and Annex 3 (part A, Section 5) of Directive 91/414. Guidance for generating and reporting methods of analysis. SANCO/3029/99 rev.4. 11/07/00

European Commission, 2002. Guidance Document on Terrestrial Ecotoxicology Under Council Directive 91/414/EEC. SANCO/10329/2002-rev. 2 final, 17 October 2002

European Commission, 2003. Guidance Document for environmental risk assessments of active substances used on rice in the EU for annex I inclusion. Final Report of the Working Group "MED-RICE" prepared for the European Commission in the framework of Council Directive 91/414/EEC. SANCO/1090/2000-rev. 1, June 2003

European Commission, 2011. Guidelines on comparability, extrapolation, group tolerances and data requirements for setting MRLs. SANCO 7525/VI/95-rev. 9. March 2011. p. 1–46.

European Commission, 2012. Guidance document on the assessment of the equivalence of technical materials of substances regulated under Regulation (EC) No 1107/2009. SANCO/10597/2003-rev. 10.1, 13 July 2012.

European Commission, 2013. Working document on the nature of pesticide residues in fish. SANCO/11187/2013 rev. 3. 31 January 2013.

European Commission, 2017. Appendix D. Guidelines on comparability, extrapolation, group tolerances and data requirements for setting MRLs. 7525/VI/95-rev.10.3

European Commission, 2018. Technical guidelines for determining the magnitude of pesticide residues in honey and setting Maximum Residue Levels in honey. SANTE/11956/2016 rev. 9, 14 September 2018.

FOCUS (Forum for the Co-ordination of Pesticide Fate Models and their Use), 2006. Guidance document on estimating persistence and degradation kinetics from environmental fate studies on pesticides in EU Registration Report of the FOCUS Work Group on Degradation Kinetics. EC Document Reference SANCO/10058/2005-v. 2.0, 434 pp.

Gonsior G and Schwalbach N, 2014. *Myriophyllum* Bio-TestSETAC Europe 24th Annual Meeting Abstract Book.

ISO (International Organisation for Standardization), 1995. Soil quality: Determination of organic and total carbon after dry combustion (elementary analysis) (10694:1995). Revised in 2016.

ISO (International Organisation for Standardization), 1998. Soil quality: Determination of the water-retention characteristic -- Laboratory methods (ISO/DIS 11274).

ISO (International Organisation for Standardization), 1999. Soil quality: effects of pollutants on earthworms. Part 3: guidance on the determination of effects in field situations (ISO 11268-3). Revised in 2014. 1–8.

ISO (International Organisation for Standardization), 2005. Soil quality – Determination of pH (10390:2005). Revised in 2015.

ISO (International Organisation for Standardization), 2006. Soil quality – Sampling of soil invertebrates – Part 1: Hand-sorting and formalin extraction of earthworms (ISO 23611-1:2006).

ISO (International Organisation for Standardization), 2009. Soil quality: Determination of particle size distribution in mineral soil material – Method by sieving and sedimentation (11277:2009). Revised in 2015.

ISO (International Organisation for Standardization), 2014. Specifies techniques for determining the effects of substances on earthworms in the field and provides a basis for determining the effects of chemicals applied to or incorporated into soil, including soil injections or drilled pelleted formulations (ISO 11268-3: 2014)

Lenat D, 1988. Water Quality Assessment of Streams Using a Qualitative Collection Method for Benthic Macroinvertebrates. Journal of the North American Benthological Society, 7, 222-233. doi:10.2307/1467422

Ludwigs J, Ebeling M, Fredricks TB, Murfitt RC and Kragten S, 2017. Appropriate exposure estimates for wildlife risk assessments of crop protection products based on continuous radio telemetry: A case study with woodpigeons. Environmental Toxicology and Chemistry, 36, 1270-1277. doi:10.1002/etc.3656

Mead-Briggs MA, Moll M, Grimm C, Schuld M, Ufer A and Walker H, 2010. An extended laboratory test for evaluating the effects of plant protection products on the parasitic wasp, Aphidius rhopalosiphi (Hymenoptera, Braconidae). BioControl 55(3), 329-338.

NL, 2017 Evaluation Manual for the Authorisation of Plant protection products according to Regulation (EC) No 1107/2009. Chapter 7 Ecotoxicology; aquatic; version 2.2; April 2017.

OECD (Organisation for Economic Co-operation and Development), 2004a. Test No. 218: Sediment-Water Chironomid Toxicity Using Spiked Sediment, OECD Guidelines for the Testing of Chemicals, Section 2, OECD Publishing, Paris. doi:10.1787/9789264070264-en

OECD (Organisation for Economic Co-operation and Development), 2004b. Test No. 219: Sediment-Water Chironomid Toxicity Using Spiked Water, OECD Guidelines for the Testing of Chemicals, Section 2, OECD Publishing, Paris. doi:10.1787/9789264070288-en

OECD (Organisation for Economic Co-operation and Development), 2006a, Test No. 208: Terrestrial Plant Test: Seedling Emergence and Seedling Growth Test, OECD Guidelines for the Testing of Chemicals, Section 2, OECD Publishing, Paris, https://doi.org/10.1787/9789264070066-en.

OECD (Organisation for Economic Co-operation and Development), 2006b, Test No. 227: Terrestrial Plant Test: Vegetative Vigour Test, OECD Guidelines for the Testing of Chemicals, Section 2, OECD Publishing, Paris, https://doi.org/10.1787/9789264067295-en.

OECD (Organisation for Economic Co-operation and Development), 2007a. Test No. 501: Metabolism in Crops, OECD Guidelines for the Testing of Chemicals, Section 5, OECD Publishing, Paris. doi:10.1787/9789264061835-en

OECD (Organisation for Economic Co-operation and Development), 2007b. Test No. 502: Metabolism in Rotational Crops, OECD Guidelines for the Testing of Chemicals, Section 5, OECD Publishing, Paris. doi:10.1787/9789264061859-en

OECD (Organisation for Economic Co-operation and Development), 2007c, Test No. 503: Metabolism in Livestock, OECD Guidelines for the Testing of Chemicals, Section 5, OECD Publishing, Paris. doi:10.1787/9789264061873-en

OECD (Organisation for Economic Co-operation and Development), 2007d, Test No. 504: Residues in Rotational Crops (Limited Field Studies), OECD Guidelines for the Testing of Chemicals, Section 5, OECD Publishing, Paris, https://doi.org/10.1787/9789264013384-en.

OECD (Organisation for Economic Co-operation and Development), 2007e, Test No. 505: Residues in Livestock, OECD Guidelines for the Testing of Chemicals, Section 5, OECD Publishing, Paris. doi:10.1787/9789264061903-en

OECD (Organisation for Economic Co-operation and Development), 2009. Test No. 509: Crop Field Trial, OECD Guidelines for the Testing of Chemicals, Section 5, OECD Publishing, Paris. https://doi.org/10.1787/20745796

OECD (Organisation for Economic Co-operation and Development, 2012. Test No. 305: Bioaccumulation in Fish: Aqueous and Dietary Exposure, OECD Guidelines for the Testing of Chemicals, Section 3, OECD Publishing, Paris, https://doi.org/10.1787/9789264185296-en.

OECD (Organisation for Economic Co-operation and Development), 2013. Guidance document on residues in livestock. In: Series on Pesticides No 73. ENV/JM/MONO(2013)8, 04 September 2013/9789264076457-en

OECD (Organisation for Economic Co-operation and Development), 2014a. Test No. 238: Sediment-Free Myriophillum Spicatum Toxicity test, OECD Guidelines for the Testing of Chemicals, Section 2, OECD Publishing, Paris. doi:10.1787/9789264224131-en

OECD (Organisation for Economic Co-operation and Development), 2014b. Test No. 239: Water-Sediment Myriophyllum Spicatum Toxicity Test, OECD Guidelines for the Testing of Chemicals, Section 2, OECD Publishing, Paris. doi:10.1787/9789264224155-en

OECD (Organisation for Economic Co-operation and Development), 2016. Guidance Document for Conducting Pesticide Terrestrial Field Dissipation Studies. Series on Testing & Assessment No 232. Series on Pesticides No 82. OECD Guidelines for the Testing of Chemicals. OECD Publishing, Paris.

OECD (Organisation for Economic Co-operation and Development), 2018. Guidance Document On Aquatic Toxicity Testing Of Difficult Substances And Mixtures ENV/JM/MONO(2000)6/, 6 July 2018.

Römbke J, Kula C and Riepert F, 2006. Technical Recommendations for the Update of the ISO Earthworm Field Test Guideline (ISO 11268-3). Journal of Soils and Sediment, 6, 182-186.

South Member States, 2018. Working document on the work-sharing of the southern zone member states under Regulation EC 1107/2009. Rev. 7.1 of 06.08.2018.

Van Capelle C, Schrader S and Arpaia S, 2016. Selection of focal earthworm species as non-target soil organisms for environmental risk assessment of genetically modified plants. Science of the Total Environment, 548-549, 360-369.

van Wijngaarden RP and Arts GH, 2018. Is the tier-1 effect assessment for herbicides protective for aquatic algae and vascular plant communities?. Environmental Toxicology and Chemistry, 37, 175-183. doi:10.1002/etc.3936

## Abbreviations

| | |
|---|---|
| AGD | aquatic guidance document (EFSA PPR Panel, 2013) |
| AF | assessment factor |
| ANOVA | analysis of variance |
| a.s. | active substance |
| AUC | area under the curve |
| BBCH | Biologische Bundesanstalt Bundessortenamt und CHemical Industrie |
| DAR | draft assessment report |
| DFOP | double first-order in parallel dissipation kinetic |
| $DT_{50/90}$ | period required for 50/90 % dissipation (define method of estimation) |
| $EC_X$ | effective concentration at X per cent |
| $Eb/yC_{50}$ | effective concentration (biomass/yield) |
| EPT | Ephemeroptera, Plecoptera, Trichoptera (orders of insects) |
| $ErC_X$ | effective concentration (growth rate) at X per cent |
| $EbR_{50}$ | effective rate (biomass) |
| ERO | ecological recovery option |
| ESI | electrospray ionisation |
| ETO | ecological threshold option |
| EU | European Union |
| $f_{TWA}$ | time-weighted average factor |
| FOCUS | Forum for the Co-ordination of Pesticide Fate Models and their Use |
| FOMC | first-order multi-compartment dissipation kinetics |
| GAP | good agricultural practice |
| $HC_5$ | hazard concentration |
| HPLC-UVD | high performance liquid chromatography with ultra-violet detector |
| HS | Hockey Stick dissipation kinetic |
| $LC_X$ | Lethal concentration at X per cent |
| LC–MS/MS | liquid chromatography with tandem mass spectrometry |
| LOD | Limit of detection |
| LOQ | limit of quantification |
| MDD | minimal detectable difference |
| MRL | maximum residue level |
| MRM | multiple reaction monitoring |
| NOAEC | no observed adverse effect concentration |
| NEU | northern EU |
| NOEC | no observed effect concentration |
| NTA | non-target arthropods |
| NW | normalised width |

| | |
|---|---|
| OECD | Organisation for Economic Co-operation and Development |
| PBI | plant-back interval |
| PEC | predicted environmental concentrations |
| PPP | plant protection products |
| PPR Panel | EFSA Panel on Plant Protection Products and their Residues |
| PT | proportion of daily diet obtained in the treated area |
| $P_{ow}$ | partition coefficient between n-octanol and water |
| QuEChERS | quick, easy, cheap, effective, rugged, and safe (analytical method) |
| RAC | regulatory acceptable concentration |
| RAR | renewal assessment report |
| RMS | rapporteur Member State |
| RSD | relative standard deviation |
| RUD | residue per unit dose |
| SEU | southern EU |
| SFO | single first-order dissipation kinetic |
| SSD | species sensitivity distribution |
| TG | test guidelines |
| TRR | total radioactive residues |
| TWA | time-weighted average |
| VDF | vegetation distribution factor |

# Appendix A – Examples of fit-for-purpose analytical methods

## Example 1

Validation of a method for the determination of substance A in water.

**Principle of the method**
The water samples were enriched and further saturated with acetonitrile and water. The final determination of substance A was performed with HPLC-UVD (λ: 220 nm).

### A. Materials

Standards:

Reference items for fortification and calibration:
Substance A
Lot/Batch no.: xxx
Purity: 99.4 %

Matrix:

Reconstituted water including analytical grade salts

### B. Control specimen, recoveries and analytical calibration

Recoveries:

Fortification level: 25 mg/L (limit of quantification, LOQ), 50 mg/L (2 × LOQ) and 100 mg/L (4 × LOQ)
Sample no.: 4 per fortification level

Analytical calibration: External standard calibration
Control matrix: Reconstituted water

**Results**

**A. Linearity**

Calibration was performed by external standard at seven concentrations ranging from 0.5 to 12.5 mg/L.
For the analysis of the fortified samples, solutions were diluted by a factor of 10 to fit within the above-stated calibration range. The resulting test substance peak areas versus theoretical test substance concentration data were fitted into a linear function.

A typical equation was determined to be:

| Calibration range [mg/L]: 0.5 – 12.5 | Equation: $y = 458482x - 10119$<br><br>where $y$ is the response in the chromatogram and $x$ the concentration of the substance [mg/L] | Correlation coefficient: $r = 0.9999$ |
|---|---|---|

**B. Specificity**

No interference (< 30 % LOQ) of total peak area for the target analyte was found in unfortified control samples.

**C. Limit of quantification**

The LOQ (corresponding to the lowest fortification level in this study) for the determination of substance A was determined at 25 mg/L.

**D. Precision (repeatability, validation in the same laboratory) and recovery**
Recovery results for the determination of substance A in water are reported in Table A1.

**Table A1:** Recovery results for the determination of substance A in water.

| Matrix | Fortification level [mg/L] | n | Range | Recovery rates [%] mean ± RSD | Overall mean ± RSD |
|---|---|---|---|---|---|
| water | 25 | 4 | 111–114 | 112±2 | 112 ± 2 |
| water | 50 | 4 | 111–116 | 113±3 | |
| water | 100 | 4 | 109–115 | 112±3 | |

RSD = relative standard deviation

**Conclusion**

The recovery is outside the acceptable range of 70–110 % according to SANCO/3029/99 rev.4 (European Commission, 2000). The slight exceedance of 2–3 % is accepted: fit for purpose.

**Example 2**

Validation of an analytical method for the determination of substance B in the feeding stock solutions and the feeding solutions used in support in the ecotoxicological study on bees.

**Principle of the method**
The stock solution samples were diluted into two different media:

- for Diet B: 15% (w/v) glucose, 15% (w/v) fructose, 3% (w/v) yeast
- for Diet C: 15% (w/v) glucose, 15% (w/v) fructose, 3% (w/v) yeast

Consecutively, the feeding solution samples were diluted in 50/50 (w/w) royal jelly / stock solution from Diet B or C.

A known weight of Diet C was spiked with substance B. An aliquot of 1 g was then analysed following a QuEChERS extraction procedure. After shaking and centrifugation, the acetonitrile-QuEChERS extracts were diluted with a mixture of acetonitrile/water (50/50) and the determination was conducted by an in-house developed method using reverse – high performance liquid chromatographic (HPLC) detection.

Detector: ESI positive, MRM: m/z 324/262, 323/242

Validation data presented for one transition.

**Conclusion**
The analytical method described in this study is considered to be fit for purpose for the determination of substance B in test item feeding solutions at the relevant concentrations. To fully comply with the requirements, the confirmatory transition should have been validated.

**Example 3**

Validation of a method for the determination of substance C in water. The method was used in support of an acute toxicity study on sheepshead minnow (*Cyprinodon variegatus)* under static conditions.

**Principle of the method**

Test solutions from the study were extracted by shaking with dichloromethane. The dichloromethane was then removed under reduced pressure and the resulting extract taken up into acetonitrile:water (1:1) prior to analysis using HPLC with UV detection (HPLC-UV, 254 nm).

During the method validation, dilution water was spiked (7 times) with substance C at different concentrations, extracted and analysed (see Table A2).

**Table A2:** Summary of validation data for substance C in salt water (sheepshead minnow media)

| **Matrix** | **LOQ (µg/L)** | **Fortification level (µg/L)** | **Recoveries % range (mean)** | **% RSD (n)** | **Linearity** |
|---|---|---|---|---|---|
| Salt water | 10 | 10<br>100<br>400<br>1600 | 90–94 (92.33)<br>86–92 (89)<br>93<br>103 | 2.3 (3)<br>- (2)<br>-<br>- | No information provided |
| Overall mean = 93<br>Overall %RSD = 5.16 (7) | | | | | |

LOQ: limit of quantification; RSD = relative standard deviation

**Conclusion**

Appropriate example LC–MS/MS chromatograms were provided which illustrate that there are no significant interferences. The method is considered suitably specific for the analytes of interest.

No information pertaining to the linearity is provided. Therefore, it cannot be concluded whether the method is validated in accordance with SANCO/3029/99 rev. 4. (European Commission, 2000) for the determination of the concentration of substance C in salt water. However, the method is considered fit for purpose.

# Appendix B – Use of residue data to support ecotoxicological assessments

## Standard residue studies that may be useful in supporting the ecotoxicological risk assessment of pesticides

### B1. Overview of the studies available in the residue data package

This appendix provides an overview of the available studies in the residues section (Vol. 3 B.7.) of the assessment reports (DAR/RAR) that might be used to support the ecotoxicological assessment of active substances.

The studies listed below are required under the European data requirements for PPPs (Regulation No. 283/2013):

i. Metabolism studies in primary crops
ii. Metabolism studies in rotational crops
iii. Supervised residue trials
iv. Metabolism studies in livestock
v. Feeding studies
vi. Other studies (residues in pollen and bee products).

An overview of the studies mentioned above, including the purpose and some relevant parameters for environmental risk assessors' consideration is available in Table B1.

Even though the interpretation of the results must be confirmed by a pesticide residue specialist, given that experts' judgement is deemed necessary in many risk assessment sections, Section 2 of the appendix provides certain considerations for assessing the residue information when relevant for non-target taxa in the field.

**Table B1.** List of the studies that are part of the residue data package (RAR/DAR Vol. 3 B.7) that might give indications of the residue situation to support the ecotoxicological assessment of pesticides (RAR/DAR Vol. 3 B.9).

| Information available in Vol. 3 B.7 | Metabolism in primary crops (TG 501: OECD, 2007a) | Metabolism in rotational crops (TG 502: OECD, 2007b) | Decline supervised residue trials (TG 509: OECD, 2009) |
|---|---|---|---|
| **Purpose of the study** | - Elucidate the metabolic pathway of the active ingredient in treated crops.<br>- Identify the major components of the terminal residue in the edible parts of the crop.<br>- Provide an estimate of the total residues in the various raw agricultural commodities after the crop treatment, which allows the determination of the distribution of the residues within the crop, e.g. whether the pesticide is absorbed though roots or foliage or whether translocation occurs. | - To assess the potential for the pesticide and its soil metabolites to accumulate in a rotational/succeeding crop.<br>- Studies particularly relevant for soil metabolite uptake. | - Determination of the magnitude of the residues under realistic field conditions according to the pertinent residue definitions over time. |
| **Considerations for the use of studies in environmental risk assessment** | - The study design and the use pattern under assessment should be comparable in terms of application rate, BBCH, number of applications, sampling, etc.<br>- Primary studies should reflect the intended use pattern of the active substance as foliar, soil/seed treatment and the post-harvest treatments.<br>- Whenever a unique situation is foreseen due to specific growth conditions, specific metabolism studies are required, i.e. paddy rice, genetically modified crops need to be supported by specific metabolism studies.<br>- Crop groups are classified under the categories specified in OECD 501 and extrapolations are foreseen in Annex 1.<br>- Residue data in the different parts of the primary crop and over time need to be available, i.e. characterisation/identification of radioactive residues and their distribution expressed as both percentage of the total radioactive residue (% TRR) and concentration (mg/kg).<br>- Representative sampling, e.g. samples in edible and inedible parts of the crop might be available. Particular attention must be given to the BBCH at sampling time. | - The maximum seasonal application rate in the rotational crop study must cover the GAPs under assessment.<br>- Rotational crops should be representative of each of the following crop groupings: root and tuber vegetable, e.g. radish, beets or carrots; small grain, e.g. wheat, barley, oats or rye; and leafy vegetables, e.g. spinach or lettuce.<br>- Studies should be performed at an application rate equivalent to the maximum seasonal rate.<br>- Quantification of the residues expressed as mg/kg and %TRR in different crop products (mature/immature lettuce, wheat straw, etc.) and at different plant-back intervals (PBIs) (at least three different PBIs) needs to be available. | - Supervised residue trials should be GAP-compliant in terms of the number of applications, BBCH of the applications and sampling and application rate(s).<br>- Determination of residues should be performed at different times, expressed as mg/kg.<br>- Independence among residue trials should be demonstrated. If residue trials are considered replicates, the selection of the residue values should be checked according to EFSA, 2015. |

| Information available in Vol. 3 B.7 | Metabolism in livestock[5] (TG 503: OECD, 2007c) | Feeding studies in livestock (TG 505: OECD, 2007e) | Residue level in pollen and bee products (European Commission, 2018)[6] |
|---|---|---|---|
| **Purpose of the study** | - Elucidate the metabolic pathway for pesticides in ruminants and poultry.<br>- Identify the major components of the terminal residue in the edible tissues, thus indicating the components to be analysed in residue quantification studies.<br>- Provide an estimate of the total residues in the edible livestock commodities as well as in the excreta.<br>- Provide evidence as to whether a pesticide residue can be classified as fat soluble. | - Establish the maximum residue levels for consumers' protection.<br>- Determination of the residues in the different animal tissues after the exposure to three different feeding levels. I.e. feeding is provided by capsules that contain the expected residue concentrations in feed and consistent exposure over the duration of the study. | - Determination of the magnitude of the residues in pollen, nectar and/or aerial parts of the plant following the GAP under assessment. |
| **Considerations for the use of studies in environmental risk assessment** | - The ruminant metabolism studies are carried out with one single animal, while poultry studies comprise 10 hens per experiment.<br>- Control animals are not necessary.<br>- Experiments are carried out with radiolabelled material and considering overdosing conditions. This is the reason why metabolism studies are qualitative and not quantitative.<br>- The test item should be representative of the residue situation of the crop.<br>- Pesticide residues in the different part of the animal tissues are determined at sacrifice while determinations in eggs and milk are of daily sampling to establish the plateau concentration.<br>- Residue determinations in the different parts of the animal tissues at the different dose levels and over time need to be available, i.e. characterisation/identification of radioactive residues and their distribution expressed as both percentage of the total radioactive residue (% TRR) and concentration (mg/kg).<br>- It must be noted that the metabolic pathway in rodents (typically rats) might be different from that in ruminants; thus, risk assessors should carefully check the appropriateness of the studies for extracting results. | - Residue determinations in the different parts of the animal tissues at the different dose levels and over time according to the characterisation made in the metabolism studies need to be available. I.e. residue distribution in different tissues expressed as both percentage of the total radioactive residue (% TRR) and concentration (mg/kg).<br>- This information might be used in a qualitative way to understand whether there is a possible accumulation in animal tissues over time.<br>- It must be noted that the test guidelines indicate that extrapolation to other domesticated animals is appropriate. Nevertheless, further consideration of a proper comparability of the studies might be carefully evaluated by risk assessors. | - Supervised field residue trials should be GAP-compliant in terms of number of applications, BBCH of the applications and sampling and application rate(s).<br>- Studies can be performed in tunnels or in the field.<br>- For setting maximum reside levels, the acceptability of the studies must be checked in line with European Commission (2018).<br>- However, further alignment with the EFSA bee guidance (EFSA, 2013) might be necessary for environmental considerations.<br>- Determination of residues should be performed at different times, expressed as mg/kg. |

[5] For fish, an update of the available guidelines is ongoing. Although not yet finalised, for further information please the last version available (European Commission, 2013).

[6] The technical guidelines for determining the magnitude of pesticide residues in honey and setting of maximum residue levels in honey published in September 2018 (European Commission, 2018), implemented by 01/01/2020.

## B2. Considerations when assessing the residue data package for environmental purposes

### B2.1. Residues in plants

The metabolic patterns of residues in plants are determined by studies in primary crops and studies in rotational crops[7]. Metabolism studies in primary crops aim to elucidate the degradation pathway of the active substance in a plant when a pesticide is applied to a crop directly (e.g. spray application) or indirectly (e.g. soil application before planting) and provide an estimate of the total residues in the various raw agricultural commodities after crop treatment.

In these studies, the identification and characterisation of plant metabolites is determined by quantification of the residues in the different parts of the crop. These data (normally expressed as both %TRR and mg/kg) do not always provide a realistic situation. While rotational crop studies are conducted in the field, metabolism studies in primary crops are conducted under overdosed rates and indoor controlled conditions. Under this premise, metabolism studies in primary crops must be used for a qualitative analysis of relevant metabolites formed in the plant and not for quantification purposes.

The extrapolation of the results of primary crop metabolism studies to other crops is only recommended when such crops belong to the same metabolism crop category (for further information, see OECD TG 501 Annex 1 (OECD, 2007a)). In addition, to consider this extrapolation, the study under assessment must be representative in terms of use pattern of the situation under evaluation presented in the GAP. For instance, results from a seed treatment in potatoes do not elucidate the metabolic pathway of an active substance sprayed on pome trees. Neither the use pattern nor the crops are comparable. However, if a metabolism study in apples is validated following OECD TG 501, the information can be used in a weight-of-evidence approach to derive residue definitions for the commodities under the whole fruit crop group, provided that the use pattern considered in the study and the GAP(s) under assessment are comparable.

To define when residue mobility occurs, the %TRR and absolute values in leaves, vines and roots at diverse sampling times/pre-harvest intervals are considered necessary. When there is a positive residue gradient to the newly formed parts of the plant, translocation occurs following an acropetal movement. This can be analysed by looking at the results of the metabolism studies in primary crops as well as when considering rotational crop studies.

Rotational crop studies represent a standard situation that it is usually applicable to crops that may grow in rotation (i.e. lettuce, potatoes, etc.). These studies are required when the crop under assessment may grow in rotation and the $DT_{90}$ for the parent compound/metabolites in soils is greater than 100 days. In this context, the application of the active substance is made to bare soil[8] in which different crops are planted at different times after the application (PBIs) simulating a kind of aged residue trial in-field. In the different plots, plant matrices are collected for the residue quantification at different time–plot combinations. Considering the design of these studies, it would be possible to identify whether soil metabolites occur in the aerial parts of the crop as a result of root uptake.

The above-mentioned studies provide information on the metabolic pathway of residues in plants and the nature of the final compounds to which non-target organisms feeding on plants might be exposed. Additionally, these studies might provide information on the possible movement of residues in the vascular system. Table B2 shows an overview of the metabolism studies in primary and rotational crops for a given active substance.

**Table B2**: Examples of the available metabolism studies in plants for a given pesticide

[7] OECD TG 501 and 502 (OECD, 2007a,b) provide the standard guidelines for the testing of chemicals that metabolism studies should follow in order to be considered eligible and for the interpretation of the results (see Table B1 for further information).
[8] The measurement of residues in soils is optional and data may not be always available (see TG 502 and TG 504 (OECD, 2007b, 2007d for further information on rotational crop studies).

| **Primary crops** (available studies) | **Crop groups** | **Crop(s)** | **Application(s)** | **Sampling** (DAT) |
|---|---|---|---|---|
| | Fruit crops | Grape | Foliar, 1x 100 + 2x 200g/ha | 18–19 $DAT_3$ |
| | | Pepper | Drip irrigation, 5 and 20 mg/plant | 33–97 DAT |
| | Root crops | Potato | Foliar, 3x 167 g/ha | |
| | Pulses/oilseeds | Bean | Foliar, 2x 250 g/ha | 4–29 $DAT_2$ |
| | Radiolabelled active substance: phenyl-UL-$^{14}$C<br>Sampling time expressed in DAT: Days after treatment; DATx: days after treatment No X | | | |
| **Rotational crops** (available studies) | **Crop groups** | **Crop(s)** | **Application(s)** | **PBI** (DAT) |
| | Root/tuber crops | Turnip<br>Swiss chard | Bare soil, 1x 534 g/ha | 30, 139, 280 |
| | Cereal (small grain) | Wheat | Bare soil, 1x 534 g/ha | 30, 139, 280 |
| | Radiolabelled active substance: Phenyl-UL-$^{14}$C and Pyridyl-2,6-$^{14}$C | | | |

DAT: days after treatment; PBI: plant-back interval.

To address the magnitude (quantification) of the residues to which plant/crop product consumers might be exposed, GAP-compliant supervised residue trials[9] in the field are required. These studies might be relevant for ecotoxicological assessments if decline residue trials in the field are performed. The quantification of residues by supervised residue trials might provide more realistic degradation times ($DT_{50}$) to replace default degradation times in the pertinent risk assessment scheme. Therefore, by replacing the default $DT_{50}$, the following parameters of the environmental risk assessment can be refined:

i. Multiple application factor and $f_{TWA}$ (time-weighted average factor) in the birds and mammals risk assessment
ii. $f_{TWA}$ in the risk assessment for bees
iii. Multiple application factors in the risk assessment for NTAs other than bees.

In these supervised residue trials, the residues might be determined at different pre-harvest intervals showing a residue degradation over time in the plant/plant products. Risk assessors should carefully check whether the residue situation presented in the supervised residue trials mimics the situation[10] under consideration.

As for crop products, pollen and nectar may be closely related to plant metabolism studies. According to the EU data requirements, these studies shall determine the residue in pollen and bee products[11] for human consumption, resulting from residues taken up by honeybees from crops at blossom. Other than in pollen and nectar, results of residues in honey[12] might be determined in these studies. This information, although valuable, is considered too uncertain to be used for risk assessment for honey bees.

[9] OECD TG 509 (OECD, 2009) provided the standard guidelines for field residue trial consideration.
[10] Particular attention must be paid to the determination of the residues in the food item under consideration and at appropriate BBCH.
[11] The determination of the residue level in pollen and bee products is part of the EU data requirements following Commission Regulation (EU) No 283/2013 (art 6.10.1).
[12] The residue definitions in honey might differ to those observed in plants. Although the current residue definitions for primary crops are applicable to honey.

### B2.2. Residues in animals

Contrary to metabolism studies in primary crops and supervised residue trials, metabolism studies in livestock are only required under certain conditions[13]. If the animal dietary burden[14] is exceeded for the use(s) under assessment, metabolism studies in livestock are triggered and the pesticide metabolic pathway in livestock needs to be identified. Common studies performed in accordance with OECD TG 503 (OECD, 2007c) include the investigation of pesticide residues in ruminants (e.g. lactating cow or goat) and poultry species (e.g. laying hens). It must be noted that these studies are used to elucidate how the residues are metabolised in animals. Thus, particular attention must be paid to the active substance under assessment as well as to the metabolites formed in plants that might be ingested by the animals. It may happen that the metabolic pathway in livestock elucidates a different residue composition than the metabolic pathway described in plants and it may also differ from the metabolic pathway observed in rodents. As for plants, metabolism studies in animals are usually performed with radiolabelled material and under overdosed conditions in order to avoid further vertebrate testing. In a qualitative way, the results provide information on the metabolic pathway of pesticide residues in livestock and determine the final residue composition in animal matrices suitable for consumption, including fatty tissues.

The final identification and magnitude of residues in products of animal origin is given by livestock feeding studies. Feeding studies are performed at three different feeding levels where animals are fed by simulating the residue concentrations in feed items and ensuring consistent exposure over time. These kinds of exposure conditions differ from the usual dietary exposure in the field. While this feeding rate is carried out with domesticated animals and in a regular regime of exposure, the field situation might not be the same and can be easily influenced by other behavioural factors (e.g. avoidance, repellence, etc.). Therefore, results from these studies must be considered carefully when assessing the exposure in wild animals.

In combination, metabolism studies and livestock feeding studies might provide information on the accumulation of pesticide residues and formation of metabolites in animals.

Although not explicitly mentioned in Table B1, fish metabolism studies and fish feeding studies are also part of the EU data requirement and the investigation of the pesticide residues when rearing fish in aquaculture conditions. Available guidelines are currently under revision; however, the same considerations as for livestock animals apply.

### B3. Conclusion

It must be noted that not all the studies mentioned in this appendix are always part of the residue section of the DAR/RAR Vol. 3 B.7 required by Regulation 283/2013. Metabolism studies in primary crops and supervised field residue trials for determining the magnitude of pesticide residues in plant and food items of plant origin are mandatory and always part of the residue data package, while other studies are triggered only under certain considerations.

The evaluation of residue data in the context of the environmental risk assessment might not be a straightforward assessment. Nevertheless, information extracted from the residue section may be considered as supportive information to build lines of evidence in order to:

- identify relevant plant and animal metabolites[15];
- either consider or disregard possible routes of exposure for non-target taxa in field conditions;
- consider available information for a more realistic field exposure;

[13] When crop products are used for feeding purposes and the expected dietary burden in livestock is greater than 0.04 mg/kg in accordance with the new EU data requirements, investigation of the residues in livestock is necessary. For further information see OECD TG 503 (OECD, 2007c) for the metabolism in livestock.
[14] The dietary burden in livestock is calculated by used the animal model 2017 that it is the excel-based implementation of OECD guidance 73 (OECD, 2013).
[15] It must be noted that animal metabolites in goats and hens might differ from the animal metabolites identified in the rat metabolism studies. Metabolism studies in rodents are available under the mammalian toxicology section (DAR/RAR Vol. 3 B.6).

- attribute to the active substance and its residues, characteristics such as fat solubility and/or the possibility of accumulation in plants or animal tissues that could not be described by using other studies.

# Appendix C – Questionnaire for the use of residue data extracted from Vol. 3 B.7. to support the ecotoxicological assessment of pesticides

The following questionnaire aims to provide a guideline for a more comprehensive characterisation of the pesticide residues that non-target organisms might be exposed to in the field. This questionnaire is of optional use and it may be considered as supporting information to be included in Vol. 3 B.9.

An empty template of the questionnaire for public use can be downloaded from the link provided. An example of how to fill it in is given below.

https://www.efsa.europa.eu/sites/default/files/corporate_publications/files/Residues_4_ecotox_questionaire_template.docx

## Metabolism in primary crops

Reference material: Test No. 501: Metabolism in Crops (OECD, 2007a)

**Question 1:** Are the provided metabolism studies in primary crops submitted in the residue section sufficient to depict a metabolic pathway of residues? If yes, which are the crop groups covered by the available metabolism studies?

Is a metabolism study available in a crop that belongs to the same metabolism crop group as the GAP(s) under assessment? Please provide an overview of the available information.[16]

**Rapporteur Member State (RMS) comment**

*E.g. three metabolism studies are presented in Vol. 3 B.7 of the DAR/RAR for active substance X. The available studies cover the use pattern under assessment in root crop group (study on potatoes) and the cereal crop group (2 studies; one in foliar application in rapeseed and one seed treatment in maize).Considering that the representative uses are on fruit (grapes) and rapeseed, only the cereals group is covered by metabolism data. A data gap for the fruit crop group has been identified.*

**Question 2:** Which are the plant metabolites recovered in the study(s) in relative and absolute amounts (greater than 10 (TRR %) and/or 0.05 mg/kg)[17] addressing the metabolic pathway of the representative use(s) [18]?

**RMS comment**

*E.g. The studies on cereals (above-mentioned) indicate that the parent compound is the most relevant compound in the extracted radioactivity and only two other metabolites were recovered at >10%TRR. The pertinent metabolites are A and B. Another three different metabolites were identified as relevant for consumer risk assessment from the potato study; however, they were not quantified in the cereal studies. It has not been possible to derive a general residue definition for the active substance.*

*The proposed residue definition for food items of plant origin is as follows: sum as parent compound and metabolite A expressed as parent compound.*

[16] The metabolism study should be conducted on a crop which belongs to the crop category representative of the GAP/intended use/representative use (e.g. a metabolism on fruit crops should be provided to support the GAP on pome fruit). It is also relevant to highlight that the metabolism study should be compliant with the GAP in terms of type of application (foliar, soil treatment, etc.), location, covering the dose rate of application, BBCH growth stage at application, pre-harvest interval.

[17] These trigger values of 0.05 mg/kg or 10 %TRR of total radioactive residues are only meant as guidance. In some circumstances, generally governed by toxicological concerns, it may be necessary to identify terminal metabolites, which are present at concentrations lower than 0.05 mg/kg or < 10 %TRR of total radioactive residues (European Commission, 1997).

[18] For the ecotox section, a selection of the relevant metabolites should reflect only the representative uses. It is not necessary to cover the residue situation for consumer risk assessment but the expected residue situation in the field for the use(s) under assessment. It is recommended to consult whether metabolism studies were summarised following harmonised templates for further assessment (I.e. EFSA/OECD templates).

**Question 3:** Is any translocation of pesticide residues observed in the different parts of the plants? Could a general conclusion be drawn on translocation of residues based on the available data? I.e. is there any particular distribution of the residues observed in specific plant tissues (leaves, grains, roots, etc.)? Is this occurring over time?[19]

**RMS comment**

*E.g. Translocation pattern in the different parts of the primary crops in which metabolism routes were investigated do not highlight a translocation pattern. Nevertheless, only information at 12 DAT was presented and residue situation overtime could not be defined.*

**Metabolism in rotational crops**

Reference material: Test No. 502: Metabolism in Rotational Crops (OECD 2007b), Test No. 504: Residues in Rotational Crops (OECD, 2007d)

**Question 4:** Do results of the rotational crops show any translocation of residues (uptake from soil) from roots to the aerial parts of the plant[20]? If so, which metabolites might be of relevance?

Is there any indication of accumulation of residues over time occurring in the rotational crop scenario? If so, in which crop categories (leafy, roots, cereals) or crop parts is the accumulation observed?

**RMS comment**

*E.g. The confined rotational crop study was conducted at an application rate of 63.9 g/ha using lettuce (leafy crop), turnips (root crop group; residues in leaves and roots) and wheat (cereal crop group) planted 30, 120 and 365 days after the application of active substance x in bare soil. Lettuce did not grow at 30, 120 or 180 days showing possible phytotoxic effects in leafy crops. Mustard has been used to replace lettuce at PBI (plant-back interval) of 300 and 365 days. Residues in green mustard (mature and immature) were found at 0.024–0.027 mg eq/kg at PBI 300 days and 0.084–0.088 eq mg/kg at PBI 365 days. Residues of active substance x were determined in turnip leaves at 0.270 mg eq/kg and 0.334 mg eq/kg at PBI 30 and 120 days, respectively, and 0.038 mg eq/kg and 0.034 mg eq/kg at PBI 30 and 120 days, respectively, in turnip roots. In wheat (forage, straw, hay and grain), the highest residues found in all crop products at 120 days PBI (0.095 mg eq/kg in wheat forage; 0.658 mg eq/kg in wheat hay; 0.555 mg eq/kg in wheat straw; 0.033 mg eq/kg in wheat grain). This study is six times overdosed in comparison with the current GAP. Nevertheless, the study highlights that residues in crops growing in rotation might account for more than 0.01 mg/kg in the equivalence to the current GAP in cereals.*

*Looking at an accumulation pattern, the study related to root crops gives an indication that there is a residue transfer from the soil to the leaves over time. Nevertheless, the same cannot be concluded for cereals grains, since a residue decline has been observed over time. For wheat forage, a residue situation has been found in the young plants (BBCH 30–40) that grow in the plots where cereals were planted 30 days after the treatment.*

**Question 5:** If the GAP is for a seed treatment or other pre-emergency[21] treatment, is any information related to the magnitude of residues at early post-emergence (BBCH < 10) for the crop(s) under assessment?

**RMS comment**

*E.g. Two out of seven GAP-compliant residue trials for cereals show a residue decline and residues were quantified at BBCH < 10. See RAR Vol. 3 B.7, trial AMI-PP-PL-2015 and trial CV-PLO9-000123.*

[19] Special attention must be given to compare results at same BBCH/sampling time; particularly for avoiding erroneous assessments due to crop growth and dissipation.

[20] It must be noted that this information may not only refer specifically to the succeeding crops/crops growing in rotation; but also, it may be useful to give indications on a possible residue situation for the new emerging plants in the crop area after certain uses. For instance, the data can be used to disregard a possible residue situation to non-target organisms originated due to the consumption of contaminated seedlings/residues in weeds.

[21] Consideration for the seedling scenario, relevant for birds and mammals and the guttation water scenario for bees might be necessary.

**Magnitude of the residues in supervised residue trial**

Reference material: Test No. 509: Crop Field Trial (OECD, 2009); Guidelines on comparability, extrapolation, group tolerances and data requirements for setting MRLs (European Commission, 2017)

**Question 6:** From the supervised residue trials, is there any indication of a residue decline over time?[22],[23] If so, please indicate the reference to the residue trial and the part of the plants where the decline was observed.

Were the residue determinations performed at 0 days after the last application or at a given time close to the last application(s)?[24]

**RMS comment**

*E.g. Residue trials were GAP-compliant and, except for the trials mentioned above, only residue information for the edible part was available.*
*For the two residue trials where residues were determined to be in decline, the parent compound was measured at 1, 3, 5, 7 and 12 DAT (days after treatment). Measurements between two applications were not carried out. Residue quantification of metabolite B included in the residue definition for risk assessment is not available. Considering that the residue samples were stored for a length of time for which the stability of the residues was demonstrated, the dataset is considered valid in terms of storage stability.*

**Question 7:** On which crops were field residue trials performed?[25] Has an extrapolation been suggested and is it considered appropriate?[26]

**RMS comment**

*E.g. GAP-compliant residue trials on wheat (northern EU (NEU) and southern EU (SEU)) and in barley (only SEU dataset) have been submitted for deriving MRLs for the representative uses in cereals (including maize, sorghum, etc.). Considering that available residue trials were validated in terms of storage stability and all were compliant with the GAPs under assessment, the validity of the residue trials is confirmed.*
*Extrapolations were considered appropriate following the EU extrapolation guidelines (European Commission, 2017) and an MRL can be derived for all crops in the group of cereals if data can be merged after the statistical analysis.*
*The assumption of similarity between NEU and SEU datasets has been statistically supported by the Mann–Whitney U-test ($p < 0.05$); however, the combined NEU-SEU dataset for barley and the combined NEU-SEU dataset for wheat were found to be statistically significantly different by the Kruskal–Williams H-test ($p > 0.05$), rejecting the possibility of merging the whole dataset.*
*Considering the information above and the existence of decline residue trials (three residue decline studies), residue information for barley and wheat can be used in combination in the same geographic region. For further information see Vol. 3 B.7 trials MM-OI22228 and IP-EE-LAO-92 and IP-EE-LAO-104-O.*

**Metabolism studies in animals (livestock, fish)**

Reference material: Test No. 503: Metabolism in Livestock (OECD, 2007c); Test No. 505: Residues in Livestock (OECD, 2007e); Test No. 305: Bioaccumulation in Fish (OECD, 2012)

[22] Please report whether the residue trials were fully validated in terms of storage stability, GAP compliance, etc.
[23] It is mentioned in the EU data requirement that when planning residue trials, it shall be borne in mind that information on the residues in ripe or unripe crops may be of interest with respect to the risk assessment in other areas like ecotoxicology and worker safety. Please include this information if available.
[24] Residue determinations close to the application(s) and/or the last application may provide relevant information for certain non-target taxa that can forage in the crop area at a time close to the application(s).
[25] The minimum number of supervised residue trials considered for setting MRLs might not be applicable for the ecotoxicology. We might build a residue decline curve with fewer than four residue data points. For this consideration, please do not disregard the residue data based only on the minimum number of residue trials. If the residue trials are compliant with GAP, ecotoxicology experts might use them for further refinements.
[26] Ecotoxicology colleagues might need advice on questions such as, 'Can residue decline studies in tomato be used to refine the residues entering the diet of frugivorous birds when the representative use is on pome trees?' and 'Can we use residue data generated in the southern EU for refinements in the northern EU zone when the representative use is in the whole EU?

**Question 8:** Is a metabolism study in fish/bioaccumulation study part of the residue section? If the fish metabolism study is available, does it indicate an accumulation of residues in fish tissues?[27]

**RMS comment**

*E.g. The metabolism study in fish is not part of the residue data package. The accumulation of the active substance in fatty tissues has been observed in other animal metabolism studies (livestock studies see below); nevertheless, the accumulation pattern observed cannot be directly linked to an accumulation of residues in fish due to different detoxification mechanisms in comparison with other vertebrate species. Pow for the active substance is 5.88 and for metabolite 1 is 5 and based on these values, residues of the parent and metabolite 1 can be preliminarily considered as fat soluble.*

**Question 9:** Can the metabolism in animals (mammals/fish/hens) bring any information on accumulation/exposure[28] to different metabolites in addition to those present in the plants?

Is it possible to observe an accumulation of residues in fatty tissues/other animal tissues considering all available metabolism studies?

**RMS comment**

*E.g. Metabolism studies in lactating goats and in laying hens are available. The metabolism pathway in goats is slightly different to the one observed in plants and to the one in rodents (mam tox section) and a new metabolite A was found in relevant amount (> 10 %TRR or > 0.01 mg/kg) in milk and liver. Available toxicological information indicates that this metabolite has a similar toxicity to the parent compound.*

*The plateau concentration in milk was not determined.*

*In laying hens, the parent compound was the major compound identified. A metabolite B was identified in < 10 %TRR in eggs. In the absence of toxicological information for this metabolite, it is not possible to conclude on its relevance for the consumers' exposure. This information has been identified as a data gap in the mam tox section; therefore, pending this information, the toxicological relevance of this metabolite might need to be further discussed.*

**Magnitude of residues in pollen and bee products**

Reference material: Technical guidelines for determining the magnitude of pesticide residues in honey and setting Maximum Residue Levels in honey (EC, 2018); Guidance on the risk assessment to plant protection products on bees (*Apis mellifera*, *bombus* spp. and solitary bees (EFSA, 2013).

**Question 10:** Are data on the magnitude of residues in pollen and bee products part of the residue section? If so, please indicate which data are available and sampling times.[29]

**RMS comment**

*E.g. Residues in pollen are part of the data package. The pollen samples were taken at the entrance of the hive. Residues were measured for the parent compound only. The study was performed in a tunnel where phacelia was treated according to the GAP under assessment. Further information was available in Vol. 3 B.7, B.7.7.1 effects on the residue level in pollen and bee products.*

*Measurements of residues in the field margin in the supervised residue trials performed according to the GAP are not available.*

*Residues in honey samples placed on the market were not available for the substance under assessment.*

[27] If we observe any accumulation in tissues, it might help in the event that further assessment of bioaccumulation and/or biomagnification (accumulation throughout trophic chain) are necessary.

[28] If there is evidence of new metabolites in the excreta, it might be relevant for the environment. Non-target organisms might be exposed to these new metabolites if there is a release in the environment after animal metabolisation.

[29] Residue section may contain information on residues in pollen, leaves and flowers. For residue assessment, data on nectar and pollen would also be useful for deriving a more realistic MRL/PF for nectar/honey and pollen/honey. Specific residue data can be used to refine higher tier studies in the risk assessment for bees if considered representative of the situation under assessment.

# Appendix D – How to present the assessment for the equivalence of batches

The equivalence of the batches used in the ecotoxicological studies to the reference source is assessed using the European Commission guidance (European Commission, 2012).

This appendix provides templates on how the information for the assessment of the compliance of the batches used in the ecotoxicological studies with the technical specification, should be presented in Volume 4 of the assessment reports. The proposed templates (see Tables D1 and D2) are intended to harmonise and facilitate the assessment of the compliance of the batches used in the ecotoxicological studies with the technical specification. For further details on how the assessment should be performed, the European Commission (2012) guidance should be consulted. Additional consideration on the assessment is also given in Section 2.5.

As agreed at the meeting and in line with the Commission guidance on the assessment of equivalence, the Tier 1 assessment should be performed and presented for all the batches used in the available ecotoxicological studies. By contrast, the Tier 2 assessment should only be conducted for the key studies, i.e. those studies that are used in the risk assessment and/or for classification purposes.

It is proposed, as a first step, to include a table reporting the list of batches used in the different ecotoxicological studies. Table D1 can be used as a template.

**Table D1:** Proposed template for the link between batch and study type

| **Batch** | **Study type** | **Author of the study and report number** |
|---|---|---|
| Batch 1 | Bird acute study | |
| Batch 2 | Fish acute study, aquatic invertebrate acute study | |
| Batch 3 | Earthworms chronic study | |
| Etc. | Etc. | |

Once all the batches are listed as proposed in Table D1, the next step is to compare their composition with the technical specification, i.e. the proposed new specification and the old specification, if any, as illustrated in Table D2.

**Table D2:** Proposed template on how the composition of batches used in the ecotoxicological studies and the technical specification should be presented.

| Short name or code of the substance | Content (g/kg) proposed in the new specification | Content (g/kg) reference specification, if any | Batch X | Batch XX | Etc. | Tier 1 assessment | Remarks[1] |
|---|---|---|---|---|---|---|---|
| Active substance | | | | | | | |
| Impurity A | | | | | | Tier 2 | e.g. Tier 2 assessment needed |
| Impurity B | | | | | | | e.g. Impurity B is a pertinent surface water metabolite. Acute data are available for fish and aquatic invertebrates showing that the metabolite is more than 100 times less toxic than the parent compound. No further assessment needed |
| Impurity C | | | | | | | Tier 2 assessment needed |

[1] The 'remarks' column is intended to include information like the need for a Tier 2 assessment or brief summary of available toxicity data (e.g. QSAR, or toxicity data which might be available in the case that one of the impurities is a pertinent metabolite).

# Appendix E – Use of $EC_{10}$ values in environmental risk assessments

Parts in yellow represent the changes with respect to Appendix F of EFSA (2015).

## 1. Introduction

During the first pesticides peer review meeting on recurring issues in ecotoxicology (Pesticide Peer Review Meeting 133, September 2015) experts from EU Member States and EFSA discussed the opportunity to request $EC_{10}$ values for toxicity tests on non-target organisms and the possible use of this estimate in the risk assessment.

The meeting concluded that for new and existing studies carried out with a suitable experimental design (which allows the calculation of $EC_X$) $EC_{10}$, $EC_{20}$, and $EC_{50}$ should be reported together with their 95 % confidence intervals. For new and existing studies where the determination of $EC_X$ is not appropriate due to the characteristics of the study design, these endpoints should not be reported, and the NOEC should be retained as a primary endpoint. In this case, it was agreed that a justification has to be provided.

In order not to repeat the same justification for studies carried out according to guidelines not optimised for deriving $EC_X$, Member States asked EFSA to compile a list of test guidelines for which the determination of $EC_{10}/EC_{20}$ should not be routinely provided. In response, EFSA scanned about 50 test guidelines and made a proposal for providing guidance on this matter (see Appendix E of EFSA, 2015).

Regarding the use of $EC_X$ in the risk assessment, the experts in the first meeting agreed that where a reliable median $EC_{10}$ could be calculated, then the lower value between this and the NOEC should be used (unless a guidance document explicitly indicates a preference; currently only EFSA PPR Panel (2013)). During this first meeting, EFSA was asked to provide guidance on the reliability assessment of $EC_{10}$. In order to fulfil this request, EFSA drew up a list of criteria to help with this evaluation (see Appendix F of EFSA, 2015).

On the basis of some further considerations made after the meeting, EFSA also proposed to consider using the 95 % lower confidence limit of the $EC_{10}$, when the median $EC_{10}$ does not provide enough certainty on the protection level. However, since neither this proposal nor the criteria given in Appendix F of EFSA (2015) had been discussed, these issues were reconsidered during the second pesticides peer review meeting on recurring issues in ecotoxicology (Pesticide Peer Review Meeting 185, October 2018). This appendix provides a synthesis of the whole process and the agreed approach.

## 2. Reliability indicators for $EC_{10}$

Any assessment of the reliability of $EC_{10}$ estimation needs the identification of suitable indicators for quantifying or, at least, comparing such 'reliability'. Two simple indicators are proposed here for this scope, both based on the concept of the confidence interval of $EC_X$.

### *2.1 Normalised width of confidence interval*

The normalised width of confidence interval (NW) is an indicator based on the relative width of the 95 % confidence interval around the $EC_{10}$ value. It is calculated as the ratio between the width of the $EC_{10}$ confidence interval and the median value of $EC_{10}$.

$$NW = \frac{(EC_{10,upp} - EC_{10,low})}{EC_{10,med}}$$

Please note that this indicator is unrelated to the shape of the dose–response curve. The relevance of this estimation for the hazard characterisation is not immediately interpretable. In principle, this indicator is applicable to any $EC_X$ estimation, not just $EC_{10}$.

### *2.2 Relationship between $EC_{10}$ and $EC_{20}/EC_{50}$ confidence intervals*

The relationship between $EC_{10}$ and $EC_{20}/EC_{50}$, unlike the NW, is very much related to the shape of the dose-response curve. The interpretation of this indicator is quite straightforward and is more related to the certainty of the level of protection ensured by the $EC_{10}$ estimation.

Two examples are analysed below.

In Figure E1, the estimated median $EC_{10}$ is lower than the lower 95 % confidence limit of $EC_{20}$. In this case it cannot be excluded that the value selected as $EC_{10}$ could result in a 'true' effect greater than 10 %, but it is very likely that this 'true' effect will be below 20 %.

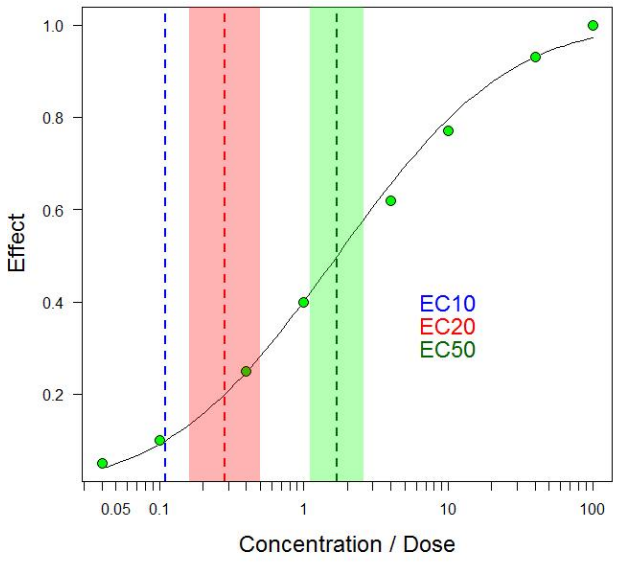



**Figure E1:** A shallow dose–response curve where the point estimate for $EC_{10}$ is clearly lower than the lower 95 % confidence limit for $EC_{20}$ (95 % confidence interval around $EC_{20}$ is shown in light red).

By contrast, in Figure E2, the estimated median $EC_{10}$ is higher than the lower confidence limit for $EC_{50}$. In this case, it cannot be excluded that the estimation of the $EC_{10}$ could result in a 'true' effect which is much higher than 10 %. Due to the uncertainty for this steep curve, it cannot be excluded that the value selected as $EC_{10}$ could in fact have 50 % effect.

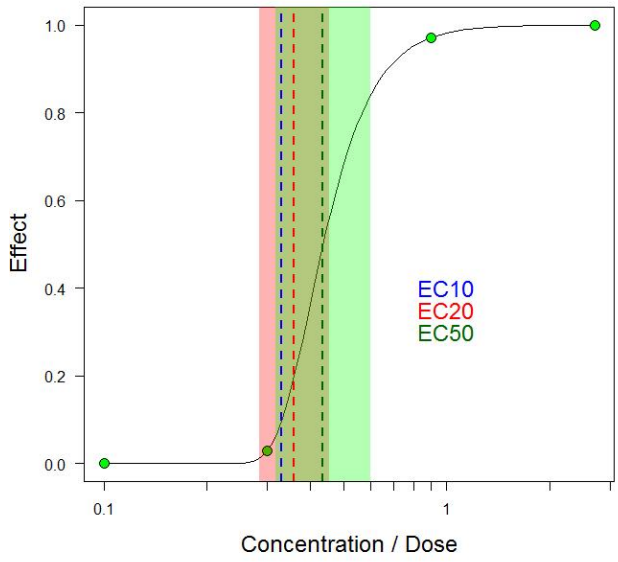



**Figure E2:** A steep dose–response curve where the point estimate for $EC_{10}$ estimation is higher than the lower 95 % confidence limit for both $EC_{20}$ and $EC_{50}$ (95 % confidence interval around $EC_{20}$ and $EC_{50}$ are shown in light red and light green, respectively).

It is worth noting that the situation described in Figure E2 is not only due to a steep dose–response, but also due to inappropriate dose spacing. A preliminary range-finding test should have highlighted that the transition between 0 % and 100 % effect was entirely completed in the 0.1–1 mg/L interval. The test should have concentrated on this concentration range. Indeed, adding another tested concentration close to the 50 % effect (see Figure E3), would considerably have improved the confidence around the $EC_{50}$ and, in turn, improved the trust that the selected $EC_{10}$ would not result in such high effects. However, this remark is mainly relevant for study designers rather than assessors.

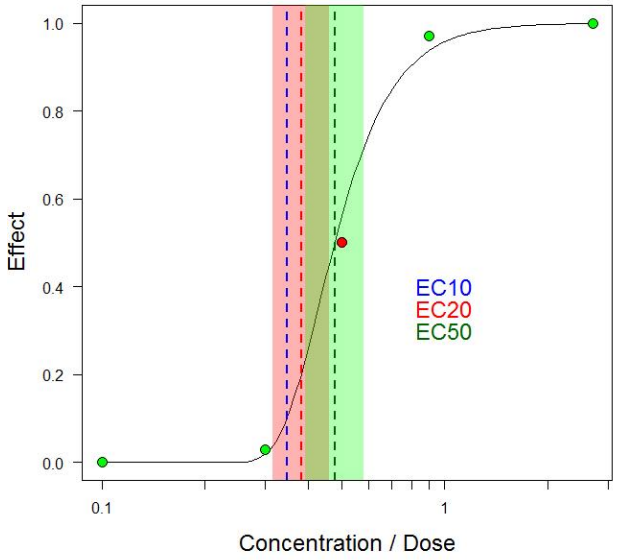



**Figure E3**: The same data as reported in Figure E2 with the addition of a tested concentration close to 50 % effect (highlighted in red)

As easily inferred from Figure E1–3, the overlap of the $EC_X$ confidence intervals is determined by two main factors: 1) the width of the confidence intervals and 2) the steepness of the dose–response curve.

It must be stressed that the former factor depends not only on the width of the confidence interval around $EC_{10}$, but also on those of $EC_{20}$ and $EC_{50}$. There might be cases where the overlap is primarily based on very wide confidence intervals for $EC_{20}$ and $EC_{50}$, even if the one for $EC_{10}$ is reasonably narrow. This is particularly true when the experimental data do not cover the entire dose–response, and the higher $EC_X$s are estimated outside of the tested concentrations. In such cases, the use of this indicator is rather meaningless, as the data may be suitable to appropriately describe the dose–response curve around 10 % effect, but not at higher effect levels.

## 3. Factors influencing the reliability of $EC_{10}$

Assessing the reliability of $EC_{10}$ is often not straightforward, and a number of issues should be considered. Factors that influence the relative width of the confidence interval around $EC_{10}$ will be discussed. Following the logic introduced in Section 2, it should be borne in mind that the same factors, together with the shape of the dose–response curve, will also affect the relationship between $EC_{10}$ and $EC_{20}$/$EC_{50}$ confidence intervals.

Multiple factors contribute to determine the relative width of the confidence interval around $EC_X$ values. These factors often have important trade-offs, so that it is quite complex to describe the influence of each single factor separately. However, a non-exhaustive list of parameters to be considered is reported below. Some simple simulations were carried out with hypothetical data sets in order to give an idea about the relevance of those factors.

### *3.1 Goodness of fit*

The most obvious factor to be checked is how well the model being used to estimate the $EC_X$ values actually describes the available data. If the data are well fitted, it is more likely that the chosen model will provide a reliable estimation of $EC_X$ values. Conversely, poor fitting would increase the uncertainty around any estimation.

It follows that when data do not describe a clear dose–response, any $EC_X$ estimation will most likely have a low reliability.

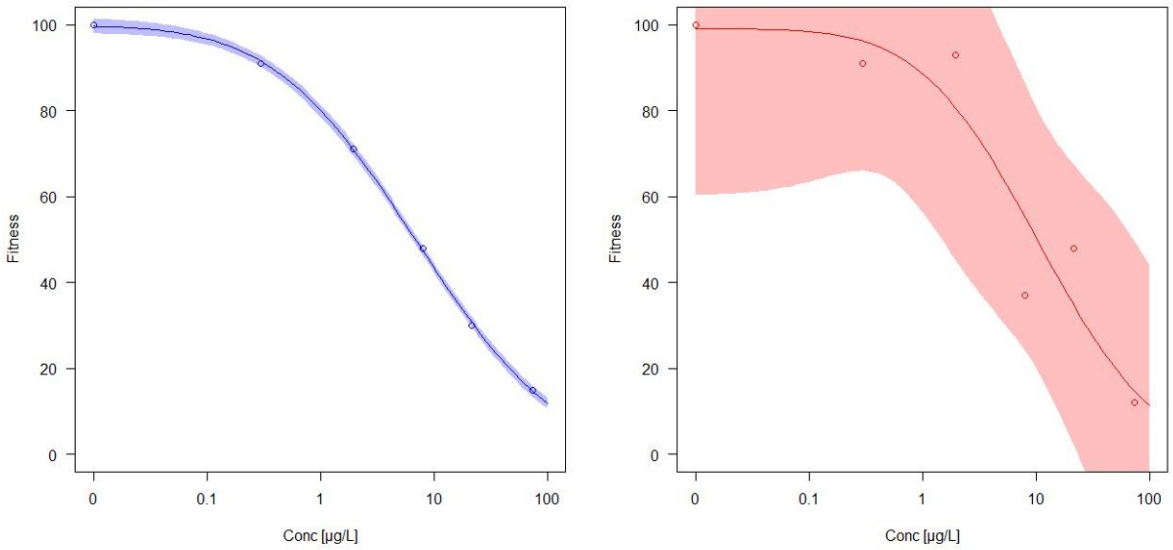



**Figure E4:** Influence of model fitting on the confidence around the dose–response curve. In the figure on the left, the curve fits the data points almost perfectly, while in the figure on the right the dose–response is hardly described

Unlike linear regression, where metrics are available to compare any situation by means of a normalised scale (e.g. $R^2$) there is no indicator to quantify the 'absolute' goodness of fit provided by a certain dose–response model.

However, some criteria are at least available to compare the performance of different dose–response models against the same dataset. These include Chi-square ($\chi^2$), Akaike's information criterion, residual standard error, lack-of-fit tests and log likelihood estimations (see, for example, the function *mselect* within the R package 'DRC').

During the second pesticides peer review meeting on recurring issues in ecotoxicology (Pesticide Peer Review Meeting 185, October 2018) it was highlighted how different dose–response models could result in different widths of the confidence interval. It is indeed possible that a model which better fits the overall dataset gives a wider $EC_{10}$ confidence interval than a less 'overall' fitting model that nonetheless better describes the surrounding data of the $EC_{10}$, thus resulting in a narrower confidence interval for this effect level. This possibility should be carefully checked, and in general, the comparison of multiple dose–response models is encouraged.

Finally, a simple 'visual check' of the fitting plot is often a very powerful tool for discriminating goodness of fit. In some cases, a visualisation of the residual plot may also help to identify bias in the data fit. Some useful indications are contained in the FOCUS guidance document on kinetics (FOCUS, 2006).

### *3.2 Number of replicates and their dispersion*

As a general rule of thumb, a higher number of replicates increases the confidence in the description of the dose–response curve, and therefore increases the reliability of any $EC_X$. However, an increasing dispersion of the replicates from the mean value could reduce the efficacy of replication in achieving a narrower confidence interval around the $EC_X$.

A simple simulation could be used to illustrate this concept. Let's assume we are evaluating body weight change of earthworms in a chronic test. Let's also assume that the test is carried out using just one replicate (10 earthworms) per tested concentration.

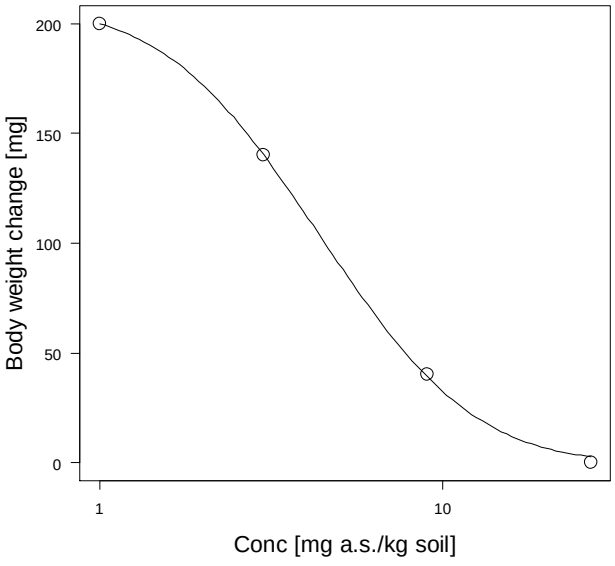



**Figure E5:** A generic dose–response curve (body weight change of earthworms)

The relevant dose–response curve is shown in Figure E5. The relevant values for $EC_{10}$ (median, lower confidence limit, upper confidence limit) are:

- $EC_{10,med}$ = 1.526 mg/kg

- $EC_{10,low}$ = 0.514 mg/kg
- $EC_{10,upp}$ = 4.525 mg/kg

The corresponding NW is: (4.525–0.514)/1.526 = 2.63 [adimensional].

To test the joint effects of replication and dispersion of the replicates from the mean values, we recalculated the NW using between 2 and 10 replicates. For the sake of simplicity, replicate values were assumed to be normally distributed around the values reported in Figure E5. Three different standard deviations were considered (3, 5, and 10). The replicate values were randomly generated in respect of the aforementioned assumptions. This procedure was repeated 1,000 times for each replication level (from 2 to 10) and for each standard deviation level, ending up with 27,000 different dose–response models (9 replication levels x 3 standard deviation levels x 1,000 iterations) and related NW values. The median NW values for each replication level and for each standard deviation level were then calculated and plotted in Figure E6.

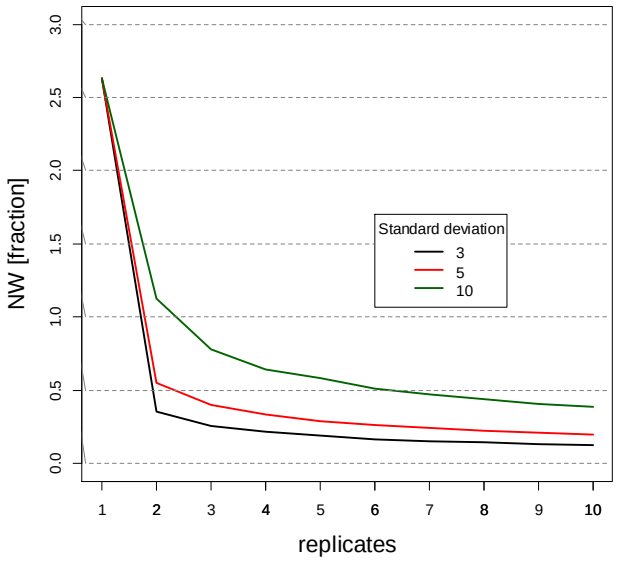



**Figure E6:** Joint effects on the normalised width due to replication and dispersion of the replicates from the mean values

Three main conclusions can be drawn from the analysis of Figure E6:

- 'Replication effect': increasing the number of replicates will increase the confidence around $EC_{10}$
- The 'replication effect' will decrease with increasing number of replicates (e.g. a much bigger decrease of NW from 1 to 2 replicates than from 9 to 10 replicates)
- 'Departure from the mean effect': increasing the standard deviation of the replicates will decrease the 'replication effect'.

### *3.3 Number of tested doses*

A high number of tested doses could greatly improve the description of the dose–response shape, as this would reduce the interval to be interpolated from one point to another. By reducing the 'guess' of the model between points, it is possible to increase the confidence around the dose–response curve.

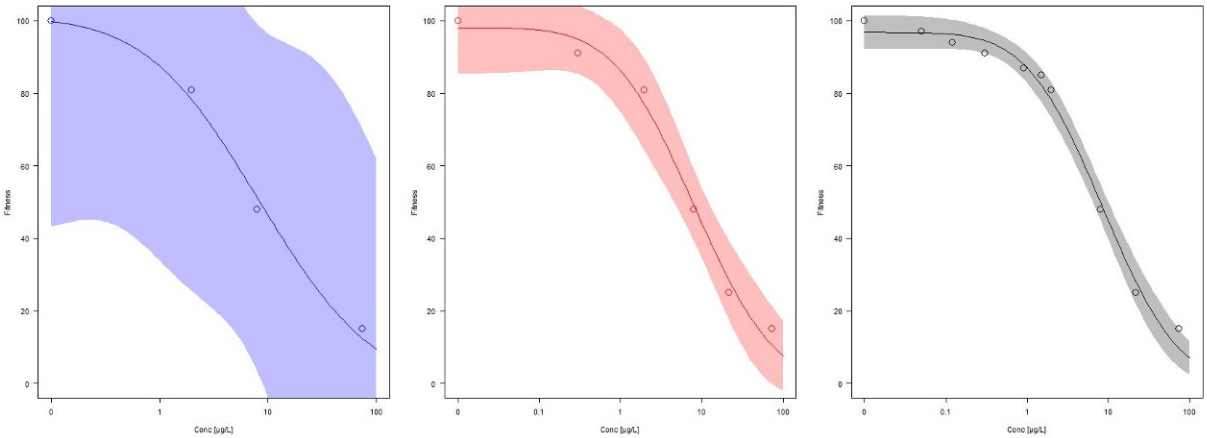



**Figure E7:** Effect of increasing the number of the tested doses on the dose–response curve confidence interval. Left figure: 4 points; middle figure: 6 points; right figure: 10 points

However, it should be considered that this is only true when further data points help in the description of the curve. Having a high number of points with 0 % or 100 % effect does not help much. For example, in the situation described in Figure E2 and Figure E3, the lowest and the highest tested concentrations do not have great importance in defining the curve shape, and have little if any effect in increasing the confidence of the most commonly estimated $EC_X$ ($EC_{10}$, $EC_{20}$, and $EC_{50}$).

In addition, the distribution of the data points around the $EC_X$ estimate of interest can make a difference. In Figure E8, both red and blue data points derive from the same response variable. For both datasets $n = 6$. However, blue points are clustered within a concentration range that yields between 0 and 20 % effects. Red points are more spaced, and the highest concentration yields up to 75 % effect. Blue points can better describe the dose–response curve between 0 % and 20 % effect. Indeed, the confidence interval around $EC_{10}$ is very narrow compared with the confidence provided by the red points. However, at higher effect levels, the confidence provided by the blue points rapidly decreases and any $EC_X$ with X > 20 is just a ventured guess which likely does not represent a reliable value (e.g. see the $EC_{50}$ illustrated in Figure E8).

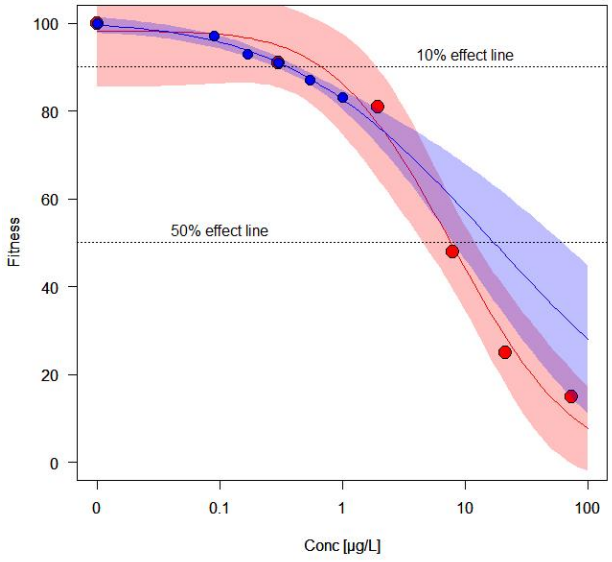



**Figure E8:** Effect of the spacing between concentrations on the confidence around the dose–response curve

### *3.4 Method of calculation of confidence interval*

Comparing the NW across different experiments is seldom a straightforward operation, and the reader should be aware of the issues that may be relevant for such a comparison. First of all, statistical software used to fit experimental data in dose–response curves has the possibility to apply different methods to estimate confidence intervals. Very often these methods rely on different error distributions, which therefore may provide non-comparable results. Further, assumptions behind dose–response fitting and confidence estimation might also be determined by the experimental data type (continuous, quantal, counts, etc.).

## 4. Practical indications from an existing database

In response to an EFSA call, a database was compiled by ICPS and Wageningen University for 'Comparison of NOEC values to $EC_{10}/EC_{20}$ values, including confidence intervals, in aquatic and terrestrial ecotoxicological risk assessment' (EFSA-Q-2013-00428). The database contains more than 800 study records on about 80 different pesticides. Details on the database can be found in the external scientific report published on the EFSA website (Azimonti et al., 2015).

With the purpose of providing general indications, the database was used to analyse the distribution of the two metrics proposed within this appendix across a wide range of different studies. On the basis of this distribution, some arbitrary ratings are also proposed. However, the reader should be aware that these ratings just have the aim of giving a very rough indication of the $EC_{10}$ reliability. Assessors are encouraged to perform a more detailed evaluation, considering all other aspects included in this appendix.

### *4.1 NW-based classification*

To implement this classification, it was considered that a NW < 0.2 should be considered as ideal. In this situation, we have 95 % confidence in saying that the true $EC_{10}$ will not be outside the estimated $EC_{10}$ ± 10 %. In the database, around 10 % of studies satisfied this condition. At the other end of the range, it was considered that when NW > 2, $EC_{10}$ estimations are likely to offer rather low reliability. In the database, this situation occurred in 12 % of cases.

Intermediate scenarios and their relative occurrence in the database are detailed in Table E9.

**Table E9:** Normalised width-based classification and occurrence of each category in the database compiled by ICPS/Wageningen University

| NW | Rating | % of cases | Cumulative % |
|---|---|---|---|
| < 0.2 | Excellent | 10.6 | 10.6 |
| 0.2-0.5 | Good | 25.9 | 36.5 |
| < 1 | Fair | 31.9 | 68.4 |
| < 2 | Poor | 19.6 | 88.0 |
| ≥ 2 | Bad | 12 | 100.0 |

NW: normalised width.

### *4.2 Classification based on the relationship between $EC_{10}$ and $EC_{20}/EC_{50}$ confidence intervals*

It has already been highlighted that this indicator is highly dependent on the shape of the dose–response curve. For this reason, it is proposed to illustrate how this classification varies on the basis of the steepness of the curve.

The steepness of the curve has been here calculated as the ratio between $EC_{10}$ and $EC_{50}$.

$$Steepness = \frac{EC_{10}}{EC_{50}}$$

A curve was classified as shallow when the steepness was < 0.33, while it was classified as steep if it was > 0.66. Every curve whose steepness was between these two thresholds was classified as medium.

$EC_X$ values normally presented (and requested by the Regulation) in ecotoxicological testing are just three: $EC_{10}$, $EC_{20}$ and $EC_{50}$. Therefore, this classification is based on the relationship between the estimated median $EC_{10}$ and its relationship with the 95 % lower confidence limit of $EC_{20}$ and $EC_{50}$. The best case (high certainty of the protection level) is achieved when the median $EC_{10}$ is lower than the lower confidence limit of $EC_{20}$. The worst case (low certainty of the protection level) occurs when the median $EC_{10}$ is greater than the lower confidence limit for $EC_{50}$.

**Table E10:** Classification based on the relationship between $EC_{10}$ and $EC_{20}/EC_{50}$ confidence intervals, considering the steepness of the curve. Occurrence of each category in the database compiled by ICPS/Wageningen University

| Condition | Rating (certainty of the protection level) | Overall | Shallow curve ($EC_{10}/EC_{50}$ < 0.33) | Medium curve (0.33 < $EC_{10}/EC_{50}$ < 0.66) | Steep curve ($EC_{10}/EC_{50}$ > 0.66) |
|---|---|---|---|---|---|
| | | **100 %** | **63.8 %** | **27.6 %** | **8.6 %** |
| $EC_{10}$ < $EC_{20,low}$ | High | 78.0 % | 91.5 % | 60.9 % | 32.8 % |
| $EC_{20,low}$ < $EC_{10}$< $EC_{50,low}$ | Medium | 14.1 % | 6.7 % | 29.3 % | 20.3 % |
| $EC_{10}$ > $EC_{50,low}$ | Low | 7.8 % | 1.8 % | 9.8 % | 46.3 % |

It has to be highlighted that in the case of steep curves, a high percentage (46.3 %) of cases within the analysed database fell within the 'low certainty' category. However, the number of steep curves in the database is very small (8.6 %).

By contrast, a very high percentage (91.5 %) of shallow curves (representing 63.8 % of all analysed curves) was in the 'high certainty' category.

It must be stressed that a very steep curve could cause a median $EC_{10}$ estimation to fall within the 'low certainty' category even if the confidence interval around such a value is not wide. In this case we cannot conclude that the $EC_{10}$ estimation is unreliable in mathematical terms; however, we cannot disregard the concerns regarding the level of protection offered by such an estimation. In a similar scenario, a reasonable approach is to take the lower limit of the 95 % confidence interval, in order to increase the confidence that the endpoint to be used in the risk assessment is not likely to cause an effect > 10 %.

## 5. Summary and final recommendations

- According to the data requirements (Regulation (EU) No 283/2013 and No 284/2013) $EC_{10}$, $EC_{20}$ and $EC_{50}$ are requested for a number of chronic/long-term tests. These values should be routinely provided for tests carried out in accordance with certain test guidelines (see Appendix E of EFSA, 2015) or upon specific request during the peer review process.
- $EC_X$ should always be reported as a median value together with the respective limits at 95 % confidence.
- It is good practice to always assess the reliability of the $EC_X$ estimation before its use in the risk assessment. Testing different dose–response models is encouraged, in order to select the one that offers the best description of the data around the relevant effect level. The reliability evaluation could consider the following criteria:
    - The width of the confidence interval around the median value;

- The certainty of the level of protection offered by the median $EC_X$, when the data allow this (e.g. for assessing $EC_{10}$, then $EC_{20}$ and/or $EC_{50}$ should not be extrapolated outside of the tested doses).

- Whenever $EC_{10}$ is to be used (e.g. median $EC_{10}$ < NOEC), but which offers scarce certainty on the actual level of protection, a reasonable approach could be to take the lower limit of the 95 % confidence interval for use in the risk assessment. The same approach can in principle be used for $EC_{50}$, although it is acknowledged that the reference given by higher effect levels is missing. Another possibility could be to select a lower effect level (e.g. 5 %), but this should be considered only in exceptional cases (e.g. $BMD_5$ for birds and mammals whenever a NOEC cannot be established).

The following scheme (Figure E) could help to clarify the overall proposal, which is based on the agreement reached in the two general meetings.

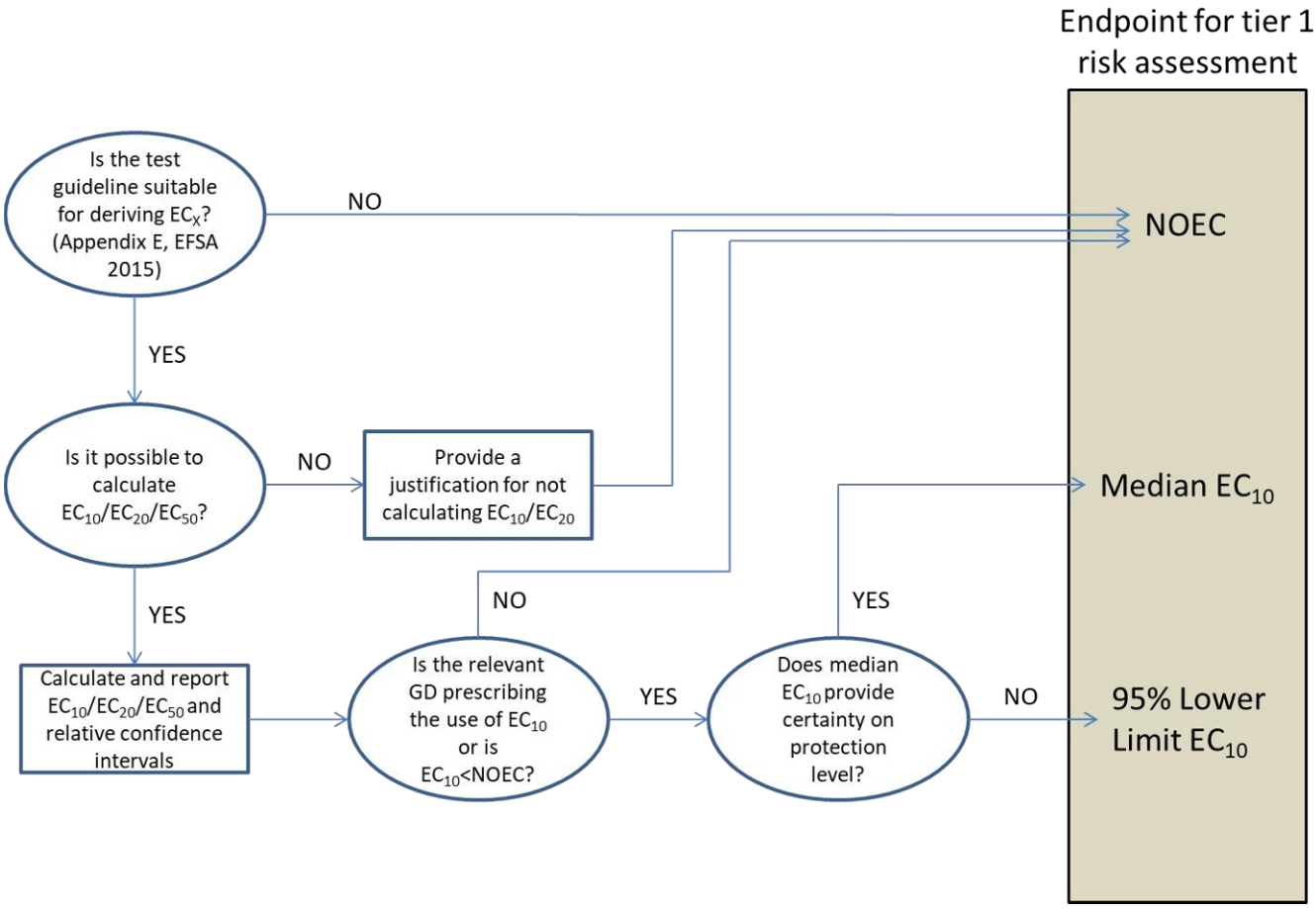



**Figure E11:** Flowchart for reporting $EC_X$ and using $EC_{10}$ vs NOEC

# Appendix F – How to assess and use trials for residue decline in the context of birds and mammals risk assessment

## 1. Definitions

*Note*

*The following definitions were not available before the meeting and have not been discussed. The same terminology may have slightly different meanings elsewhere. The aim of this section is not to come up with a universal definition of specific terms, but just to provide a basis for a common understanding of the terminology used in the present appendix.*

**Site**: one identified geographical location where a specific experiment is carried out. Sites are characterised by unique geo-climatic conditions. No clear boundaries can be set for site identification/separation. However, two sites should be considered independent if they are at sufficient geographical distance to allow some difference in the geo-climatic conditions. As a rule of thumb, ≈ 100 km is considered a sufficient distance, but in the case of very diverse landscape and topography, smaller geographical distances can still be appropriate.

**Trial**: one independent residue decline experiment providing a unique $DT_{50}$ estimation. It is characterised by a unique site, timing, target, sampling strategy and application (in terms of rate and pattern).

**Plot**: a spatially-characterised sub-unit used for a trial. Plots in the one trial are generally managed with the same experimental treatment (e.g. same crop, same applications, and same sampling strategy). They can be considered to be spatial replicates of a single trial. No general characteristics of plot dimension can be given. However, for trials designed to measure residue dissipation on arthropods a minimum of 1 ha is required for each plot (EFSA, 2009).

**Replicate**: generic term which indicates repeated measurements of the same distribution, intended to quantify the variability and the central tendency (i.e. mean, median) of that distribution. This can refer to spatial variability, temporal variability, or analytical/measurement variability.

## 2. Reliability of the single residue trials (quality of study)

| Experimental phase | General (valid for both plants and invertebrates) | Plants | Invertebrates |
|---|---|---|---|
| Design/reporting of the field phase | - Location of the experimental site(s) should be reported.<br>- Basic figures (e.g. daily min., max., mean) should be reported for temperature. Location of the weather station and distance from the trial site(s) should be available.<br>- Test item should be reported: particularly, information should be available on the type of formulation tested.<br>- Application technique, rate, timing and frequency should be reported.<br>- Sampling time, methodology and size (number of samples per site and weight) should be reported.<br>- Information on the plot handling should be reported (i.e. any relevant agronomic practice including irrigation).<br>- Sampling points should primarily cover the first few days after application(s). A sample taken before application is also recommended.<br>- Sampling at the boundaries of the plot should be avoided.<br>- Rainfall should be reported (daily values in mm). Rainfall in some cases plays a role, which should be carefully evaluated when assessing the dissipation plot. | - Plot(s) and crop characteristics should be reported.<br>- BBCH at the time of application(s) should be reported. | - Sampled taxonomic groups should be reported.<br>- Consistency of taxonomic groups should be maintained throughout the sampling phase. For this reason, sampling should be carried out at the same time of day.<br>- If there are multiple applications, it is not sufficient to start the sampling after the last application.<br>- Collection methods should be reported, and it should be evaluated whether these are relevant for foliar-dwelling/soil-dwelling organisms.<br>- At least three replicates (plots) per site should be available.<br>- Each plot should be at least 1 ha.<br>- The landscape surrounding the test area should be described.<br>- Spraying of any (additional) insecticides should be avoided. |
| Analytical phase<br>Recommendations from SANCO 3029/99 (European | - The analysed matrix should be clearly identified.<br>- The sample storage (including during transport) and stability should be reported and considered appropriate. Stability should particularly be checked | | |

| | |
|---|---|
| Commission, 2000). | when the residues are analysed more than 30 days after sampling.<br>- The analytical procedure (including sample preparation, extraction, and purification) should be summarised.<br>- Linearity, accuracy, and precision should be appropriate.<br>- LOQ and limit of detection (LOD) should be clearly reported.<br>- For invertebrates, replicates (from each plot) should be analysed separately (see Section 4 of SANCO/3029/99).<br>- It should be clearly indicated whether the results are expressed in terms of fresh or dry weight of the sample material. |

**Suggestions for study authors**

The following recommendations are not considered to influence the quality or reliability of the study, but are nevertheless highlighted as good practice to aid the assessment of the studies.

1- Rain events and their intensity should also be plotted together with results of the residue quantification over time in order to identify sudden drops in the measured residue concentrations coinciding with such events. The kinetic assessment should also consider whether such episodes have a determinant role in the dissipation of the substance.
2- It should be clearly documented by the study authors whether the weather conditions at the study plot are representative of the usual climate in the respective zone or Member State.
3- In the case of trials with crops: the weight of a defined amount of plants should be documented at different sampling points, at least the first and last point. This allows mass growth to be determined, from which the contribution of the dilution effect on total dissipation can be quantified. This is particularly relevant when assessing whether crop plants can be taken as surrogates for wild plants. Not necessary in trials performed with wild plants.

# 3. Relevance of trials for use in the risk assessment

Apart from the evaluation of their intrinsic scientific reliability, each residue study should be assessed in terms of relevance for any specific risk assessment. Below is a list of considerations that should be accounted for when assessing whether a certain residue trial can be considered for a specific risk assessment.

**Type of test**: field studies are preferred for refinement of the dissipation time. Other types of studies may be considered on a case-by-case basis if it can be shown that other aspects of the study were sufficiently worst-case to cover the uncertainty involved in the use of application apparatus other than that foreseen based on the GAP.

*Note*

*The 'type of test' paragraph was not available before the meeting and has not been discussed. The above text had been taken from the agreement reached at the Central Zone Harmonisation Meeting. Nevertheless, EFSA considers this agreement sensible and would support its inclusion in this appendix.*

**Test item**: The test item should, in principle, be the representative formulation. When this is not the case, the tested formulation should be assessed to determine whether it is sufficiently similar to the representative one (i.e. same typology). Appropriate bridging studies may also be a possibility to allow results from comparable formulations to be accepted. Attention should mainly be paid to elements which may alter the environmental fate of the active substance (e.g. encapsulation, solid vs liquid form, additives and co-formulants influencing the dissipation behaviour, etc.).

**Irrigation**: If relevant, the irrigation regime should be representative for the use under assessment.

**Application method**: The potential influence of the application method should be assessed.

**Application pattern**: Ideally, the application pattern should mimic the GAP in terms of number and frequency of applications and expected BBCH at the time of the application(s). Deviations are allowed when it could be established with reasonable certainty that the experiment is carried out under worst-case conditions (e.g. slower growth rate than the expected one in order to minimise dilution). Considering this issue, single applications under worst-case growing conditions may be considered appropriate for deriving $DT_{50}$ on vegetable material.

Arthropods: For arthropods, more flexibility should be allowed, as trials are often not carried out with the representative crop and hence a BBCH match is less meaningful. This is done with good reason, as some crops would present a poor arthropod community in terms of both diversity and abundance, and hence trials in these crops are generally not considered sufficiently informative/robust. Expert judgement should always be used to consider the representativeness of each residue trial in terms of application pattern. In addition, the number of applications can also be an important limitation. It is considered that sampling only after the last application is not appropriate, as important information regarding build-up of residues may be missed. However, sampling of arthropods can be quite resource-consuming, considering that each trial should have three plots of at least 1 ha. During the meeting, it was suggested that any build-up due to multiple applications could be addressed without conducting a test with the exact number of applications reported in the GAP (three or four could be enough). However, an exact number was not defined because this would depend on the GAP in consideration of the time-weighted average and taking into consideration available information on the dynamic and build-up of residues on arthropods.

*Note*

*EFSA does not consider the application rate (in terms of amount per hectare) to be relevant for the dissipation kinetic. However, such an indication deviates from the agreement reached at the Central Zone Harmonisation Meeting (held at the UBA in Dessau, September 2018), where the experts agreed that degradation processes may be concentration-dependent. Hence, they chose to accept trials when these are in the range of 0.3 to 4 times the proposed dose rate, following an MRL approach. In the central zone, it was also agreed that deviations larger than the ones reported above may be acceptable, but should be supported, e.g. by referring to expected processes driving*

*the residue decline and the LOQ of the substance.*

**Analysed item (matrix)**: In principle the analysed items should mirror those considered for the risk assessment. However, in general, a certain degree of extrapolation was considered possible. EFSA suggested some grouping in order to define homogeneous clusters within which extrapolation is possible. These are:

- Dicot plants (green parts and roots)
- Monocot plants (green parts and roots)
- Fruits
- Seeds (both weed seeds and cereal seeds)
- Foliar-dwelling arthropods
- Ground arthropods
- Earthworms.

Overall, the experts at the meeting agreed on the extrapolation as suggested by EFSA. However, it was pointed out that generally common sense and expert judgement should be used. Extrapolation across groups was generally not considered appropriate, but a few exceptions are possible: for example, extrapolation between dicot weeds and grass-like weeds is possible for trials performed at late growth stage. It was also agreed not to extrapolate from maize to grass-like weeds because of the fast-growing nature of maize. Residues in any other matrix (e.g. soil, water, etc.) cannot be extrapolated to any foodstuff.

## 4. Dissipation kinetics of a single trial

Dissipation kinetics should be calculated considering comparable matrices over time (e.g. sampling of whole plants and samples related to leaves cannot be considered in the same dataset for deriving dissipation kinetics). Most of the recommendations listed below are from FOCUS (2006) and are in principle valid for all sampled matrices (i.e. plants and invertebrates). Further, more specific recommendations can be retrieved therein. Worked Example A at the end of this appendix offers some more practical guidance using a fictitious case.

- At least five quantifiable time points should be available for fitting the decline curve. In some exceptional cases, four points may be enough (e.g. fast dissipation of the active substance or residues of metabolites with slow formation) but the points should never be fewer than four.
- If true replicates exist, they should be used in the fitting (averaging between replicates before fitting should be avoided). Analytical replicates, on the contrary, should always be averaged.
- A minimum number of true replicates is not required for plants, while according to EFSA (2009), three replicates should be available for arthropods. As a general rule, the reliability of the kinetic estimation increases when more replicates are available.
- For values below the LOQ/LOD, the following procedure should be followed:
  1- All values between the LOD and LOQ are set to the actual measured value. If the actual measured concentration has not been reported, use 0.5 x (LOQ + LOD).
  2- All samples < LOD are set to ½ LOD.
  3- The curve should be cut off after the pesticide has largely dissipated. All samples after the first non-detect (< LOD) should be omitted unless positive detections above LOQ are made later in the experiment. In that case, samples are included up to the first non-detect (< LOD) which is NOT followed by later positive samples above LOQ.
- If an outlier is rejected based on expert judgement, this must be clearly indicated in the report and, where possible, supported by statistical analysis. The exclusion of measured data points as outliers should also be supported by a detailed technical and scientific reasoning.
- Initial values should, as a first step, be included in the optimisation (not constrained). If a constrained procedure is to be used, this should be well justified.
- The first time point should preferably report residues sampled on the application date.
- The kinetic model (SFO, FOMC, DFOP, HS, etc.) used to fit the data should be reported, together with the relevant parameter estimates (and related 95 % uncertainty limits).
- Goodness of fit should be assessed using four indicators, all of which should be clearly reported. It should be noted that these indicators should be evaluated together and not in a hierarchical manner:

1- Visual fit → plot of time vs concentration should be provided. Ideally, the fitted line should pass through (or in the vicinity of) the measurement points.
2- Residual plot → Plot of time vs residuals against the $y = 0$ line should be provided. Points should ideally be scattered around the zero line. Regular patterns are generally indicative that the kinetic model used is not appropriate. Underestimation of the last time points is indicative of an under-conservative kinetic.
3- Chi-square ($\chi^2$) % should be reported and should ideally be < 15 %. Note that Chi-square should be calculated using the mean of true replicates.
4- *t*-test and/or confidence intervals of individual model parameters should be reported. *t*-test for rate constant resulting in *p*-values > 0.05 (or confidence intervals including zero) indicate large uncertainty in the estimation of the model parameters and such results should not be accepted.

- The software/package used to fit the data should be clearly reported.
- The selection of the appropriate kinetics (and thus of the appropriate endpoint) should follow the recommendations of FOCUS (2006) about how to derive endpoints for modelling inputs (Section 7.1.2, particularly Figure 7.2). In an extreme summary, SFO kinetics should always be preferred if the fitting is acceptable (even if other kinetics models may give a slightly better fit). Particularly for vegetable materials, SFO should always be preferred unless there is indication that the fitted model underestimates the $DT_{50}$ (not worst case). The use of pseudo $DT_{50}$ obtained with the FOMC model or of the slower $DT_{50}$ from biphasic DFOP and HS models may be appropriate if SFO cannot be used. Following the recommendation of FOCUS (2006) FOMC should not be used for sequential metabolites (fitted together with the parent).
- In general, it is recommended that for this kind of evaluation, risk assessors experienced in kinetics modelling should be involved and it may be sensible to reflect their evaluation in the DAR/RAR.

# 5. Dealing with multiple applications

*Note*

*This section was available before the meeting, but not discussed due to time constrains. No particular comments were received before the meeting. No modifications had been made since the meeting.*

In the case of residues measured after multiple applications, two alternatives are possible (Worked Example B at the end of this appendix relates to this issue). The first one is:

1- consider each application (and the following points until the next application) as a standalone trial
2- calculate as many $DT_{50}$ as the number of applications
3- calculate the geomean (see Section 6) of the calculated $DT_{50}$ as the representative for the multiple application trial.

Otherwise:

1- express all concentrations in terms of the fraction of the one measured on the last application date (i.e. on the day of each application, the value will be 1)
2- calculate the time between each measurement time point and the date of the last application
3- use the newly derived values for the fitting exercise.[30]

# 6. Use of degradation kinetics in the risk assessment

## Minimum number of trials and their combination

For residue on vegetable material, trials carried out at at least four sites per item category and regulatory zone has been proposed by EFSA as a minimum requirement in order to have a reliable refinement of the dissipation of the pesticide. It is noted that the emphasis is more on the spatial

[30] Note that this approach is likely to be successful with plant residues, while it might not be so for residues on invertebrates, where the dynamic is further complicated by other processes such as organisms' movement, different surface/body weight ratios, bioaccumulation, etc.

variability; although it is acknowledged that temporal variability may also play an important role in some cases (e.g. weather conditions for the same dates in different years may be very different).

A minimum number of trials or sites is not specified in EFSA (2009). The number suggested here had been proposed in consideration of the fact that: (a) 3–5 trials has historically been considered the minimum for refining $DT_{50}$ values; (b) the residue data requirement for MRL (minor crops) specifies that at least four independent trials should be available; (c) at least four soils should be tested for establishing a valid $DT_{50}$ in soil.

Overall, this minimum number was agreed at the meeting, although some uncertainties were pointed out. In addition, it was agreed that particular climatic conditions of certain areas allow extrapolation to some extent (e.g. northern France).

For invertebrates, EFSA proposed a reduced number of trials (i.e. one per item category and regulatory zone), in view of the practical difficulties in carrying out these studies. However, an agreed minimum number of sites/trials for invertebrates could not be established by the experts at the meeting. Hence, this is an issue to be dealt with by the ongoing working group for the revision of EFSA (2009).

As a general rule, results from different trials on equivalent items and carried out in the same regulatory zone should be averaged before being used in the risk assessment. FOCUS (2006) highlight how the geometric mean is the most solid method of averaging between studies.

*Note*

*In the report from the Central Zone Harmonisation Meeting (held at the UBA in Dessau, September 2018), the experts suggested that geometric mean can only be used in the risk assessment if it is shown that the $DT_{50}$ values for the dissipation of residues obtained from residue trials can be considered significantly different from the default value of 10 days. Further, they suggest testing this with the excel sheet provided as background information with the EFSA Guidance Document to obtain $DegT_{50}$ values (EFSA, 2014a) using a significance level (α) of 7 %.*

*EFSA does not consider this approach to be statistically sound or necessary. First, as already pointed out in the aforementioned report, the proposed comparison is performed between a mean value (the mean of the log-transformed available $DT_{50}$s) and a default value assumed to be a worst case (10 days). The proposed test checks the likelihood that the default value lies within the predicted distribution of the experimental values. However, having verified that this is the case (test accepts the null hypothesis), it does not prove that the mean value of the predicted distribution is significantly different from the default value: this claim would simply not make sense, considering the available data.*

*We are of the opinion that any sufficiently robust estimation of the actual dissipation time would be more suitable than the default value, irrespective of the relationship between the former and the latter.*

*Note*

*The following part of this section was available before the meeting, but not discussed due to time constraints. No particular comments were received before the meeting. No modifications had been made since the meeting.*

When the dissipation in all trials is described by SFO kinetics, the averaging of the $DT_{50}$ (or of the dissipation constant $k = \log(2)/DT_{50}$) is straightforward. However, if other kinetics were used, the appropriate figure to be included in the average is the following:

- FOMC (pseudo $DT_{50}$): $DT_{90}/3.32$
- DFOP and HS: slower $DT_{50}$ (unless HS: C$t$<0.9 M0 at breakpoint or DFOP: $g > 0.9 \rightarrow$ In this case use fastest $DT_{50}$).

However, the use of the worst-case value should be considered when:

- the dataset presents limited reliability (see Section 1)
- the dataset *x* is relatively small (4–6 studies) and one value is considerably higher than the others (e.g. max(*x*) - mean(*x*) > 2 * sd(*x*)).

## Splitting or merging datasets

*Note*

*This section was available before the meeting, but not discussed due to time constraints. No particular comments were received before the meeting. Following the discussion related to other parts of the document this section has been slightly amended, for the sake of consistency.*

When $DT_{50}$ values are available over multiple items and geographical areas (e.g. regulatory EU zones), a consideration should be given as to whether the $DT_{50}$ estimations are part of the same distribution. When this is the case, merging the datasets can simplify the risk assessment and provide more robust dissipation estimation[31].

This could be assessed visually (generally visualisation like boxplots helps in this case) or statistically by running the appropriate tests. When running such tests, alpha (α) levels higher than the standard 0.05 may be considered, if there is a concern that large dataset variations would lead to acceptance of the null hypothesis despite potentially influential differences in the mean values. If a higher alpha is used, this should be documented, and the underpinning reasons should be explained.

- If only two groups are present, then a *t*-test is normally the most straightforward approach. When data from the two groups clearly deviate from a normal distribution[32], then an equivalent non-parametric test can be used (Wilcoxon or Mann–Whitney tests). If the difference is significant at the chosen alpha, it is likely that the considered factor (crop, geographical area, etc.) is playing a role in the dissipation speed, and hence the two datasets should not be merged.

- If more than two groups are present (e.g. data for the three regulatory areas) an ANOVA test is the most straightforward approach. A common non-parametric equivalent for ANOVA is the Kruskall–Wallis test. Once again, if the difference between group means is found to be significant at $\alpha = 0.05$, it is likely that the considered factor is playing a role in the dissipation speed, and hence the datasets should not be merged.

- Whenever more discriminatory variables are considered to potentially have an influence (e.g. crop and regulatory areas) a two-way ANOVA (with or without interaction term) can be run. Non-parametric equivalents to two-way ANOVA exist, but they are not as straightforward as in the previous cases.

## Dealing with toxic metabolites

*Note*

*This section was available before the meeting, but not discussed due to time constraints. No particular comments were received before the meeting.*

Whenever the toxicity of the metabolites is comparable to that of the parent, the use of the TWA factor ($f_{TWA}$) based on the parent only should not be considered acceptable. In this case, the residue measurements should report both the parent and the appropriate metabolite concentrations.

[31] This procedure should not be considered a 'shortcut' for reducing the number of trials per item group and regulatory area. The minimum number of trials per item group and zone should still be respected, before reaching any conclusion on the possibility of merging datasets.

[32] Often the number of values is too small to appreciate the shape of the distribution. In this case, it is suggested to stick to the parametric test (generally more powerful).

FOCUS (2006) specifies how to fit residue data for the parent and the metabolite from the same experiment. However, one should be aware that, in this case, the number of parameters to be estimated is generally rather high. Therefore, the number of data points needed for a reliable fitting exercise should increase accordingly.

When the toxicity of the parent and the metabolite is really quite similar, an alternative approach could be to sum their concentrations at any time point and then fit the obtained data as if it was a single chemical. The derived $DT_{50}$ would then be valid for the sum of parent and metabolite. Applicative examples are presented in the Worked Example C at the end of this appendix including this and another more elaborated approach.

### Invertebrates

Residue studies with invertebrates represent a challenge for different reasons. First, unlike plants, invertebrates move in the environment, picking up the test item as they go. This bioconcentration (which may also be accompanied by dietary bioaccumulation) very often results in peak concentration being reached only a few days after the application. In addition, the concentration pattern in time is very much complicated by migrations, with sampling of animals recently arrived on the test plot, and animals initially exposed leaving the test plot and only coming back afterwards. EFSA (2009) already highlighted the fact that SFO is hardly suitable to describe the dissipation kinetics for invertebrates. With the data available at the time of the drafting of the guidance, FOMC seemed to be the best option for fitting the residues over time.

Often residue collected for repeated applications presents different patterns, with rather different initial values and dissipation speed. When the estimated $DT_{50}$ or pseudo $DT_{50}$ appears unreliable, an alternative approach could be to calculate a $f_{TWA}$ by simply integrating the area under the curve (AUC) normalised by the initial value[33] and divided by the averaging period (generally 21 days). It should be noted that this proposal differs from the use of the AUC directly in the risk assessment. The latter would mean ignoring the residue per unit dose (RUD) database given in the EFSA (2009) guidance and this is not recommended. Hence, the AUC should only be used to derive a $f_{TWA}$.

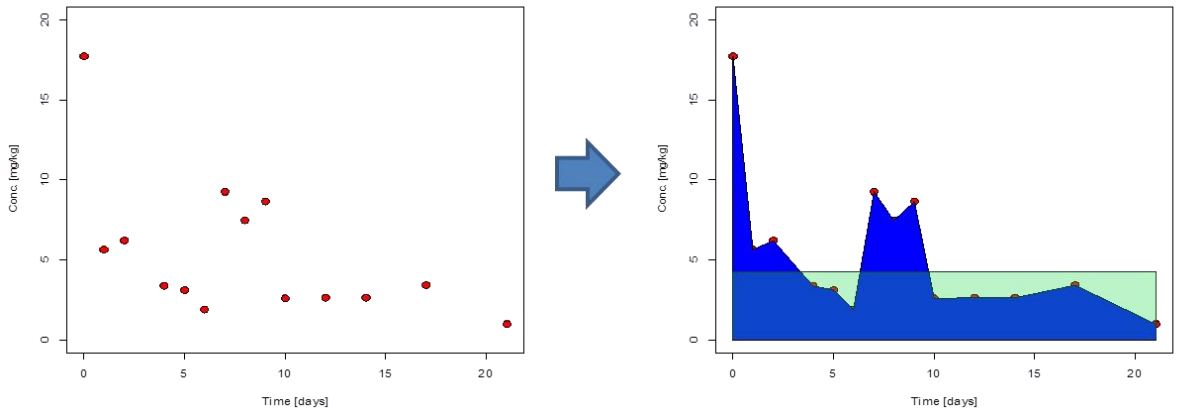



**Figure F9**: Left panel – Residues on arthropods after two applications at time = 0 days and time = 7 days. Right panel – The corresponding area under the curve is reported in blue and the equivalent area with constant concentration is reported in light green. The light green rectangle can be interpreted as the graphical equivalence of the time-weighted average concentration.

[33] At the meeting, it was questioned whether the initial or the highest value should be used for normalisation. For the sake of consistency, this should correspond to the way default RUD values were estimated in EFSA (2009). RUD values reported therein come from two separate datasets: one from ECPA and the other from CSL (Defra Project code PS2323). Only reports from the latter were found: based on the scarce information available, it seems that RUD values were based on initial measurements, as arthropods were only sampled immediately after the application (only in one case were they also sampled two days later). Hence, the use of the initial value as the normalising factor seems appropriate.

## Worked example A: Data treatment for a single trial

Here a presentation of a good dataset providing a good kinetics fitting is reported. The available fake dataset is supposed to contain residue data for a random pesticide applied once to cereals. Results are related to the whole plant.

The method of analysis was satisfactorily reported, and the LOD and LOQ are reported below:

| | |
|---|---|
| LOQ | 2 µg/kg |
| LOD | 0.66 µg/kg |

The available dataset is given in the following table.

| **DAT** | **Plot** | **Concentration [µg/kg]** | | | |
|---|---|---|---|---|---|
| | | Analysis 1 | Analysis 2 | Average analysis | Average per DAT |
| 0 | A | 10.45 | 10.82 | 10.635 | |
| 0 | B | 9.32 | 9.37 | 9.345 | 10.71 |
| 0 | C | 12.4 | 11.9 | 12.15 | |
| 1 | A | 9.2 | 9.43 | 9.315 | |
| 1 | B | 8.3 | 7.97 | 8.135 | 8.65 |
| 1 | C | 8.56 | 8.43 | 8.495 | |
| 3 | A | 7.2 | 7.45 | 7.325 | |
| 3 | B | 5.34 | 5.65 | 5.495 | 6.55 |
| 3 | C | 6.78 | 6.9 | 6.84 | |
| 7 | A | 4.2 | 4.07 | 4.135 | |
| 7 | B | 3.42 | 3.57 | 3.495 | 3.79 |
| 7 | C | 3.8 | 3.65 | 3.725 | |
| 12 | A | 1.97 | 1.99 | 1.98 | |
| 12 | B | 1.56 | 1.52 | 1.54 | 1.73 |
| 12 | C | 1.7 | 1.63 | 1.665 | |
| 21 | A | <LOD | <LOD | 0.33 | |
| 21 | B | <LOD | <LOD | 0.33 | 0.33 |
| 21 | C | <LOD | <LOD | 0.33 | |
| 42 | A | <LOD | <LOD | N/C | |
| 42 | B | <LOD | <LOD | N/C | N/C |
| 42 | C | <LOD | <LOD | N/C | |

DAT: days after treatment; LOD: limit of detection; N/C: not calculated

Note that:

- Values at 12 DAT were below the LOQ. Nevertheless, as values were still available, these were used in the fitting exercise (if they were not available, default values halfway between LOD and LOQ would have been used).
- Values at 21 DAT were all below the LOD. Therefore, they were set to half of the LOD.
- Values at 42 DAT were all below the LOD. However, the curve should be cut at the first 'non-detection' (happening in this case at DAT 21).
- Values used for the fitting are reported under the column 'Average analysis'.
    - Values in the rightmost column (average per DAT) will not be used in the dissipation fitting (single true replicates are considered independent).
    - Values in the columns 'Analysis 1' and 'Analysis 2' represent analytical replicates of the same sample, and they are averaged before the fitting.

*SFO model fit*

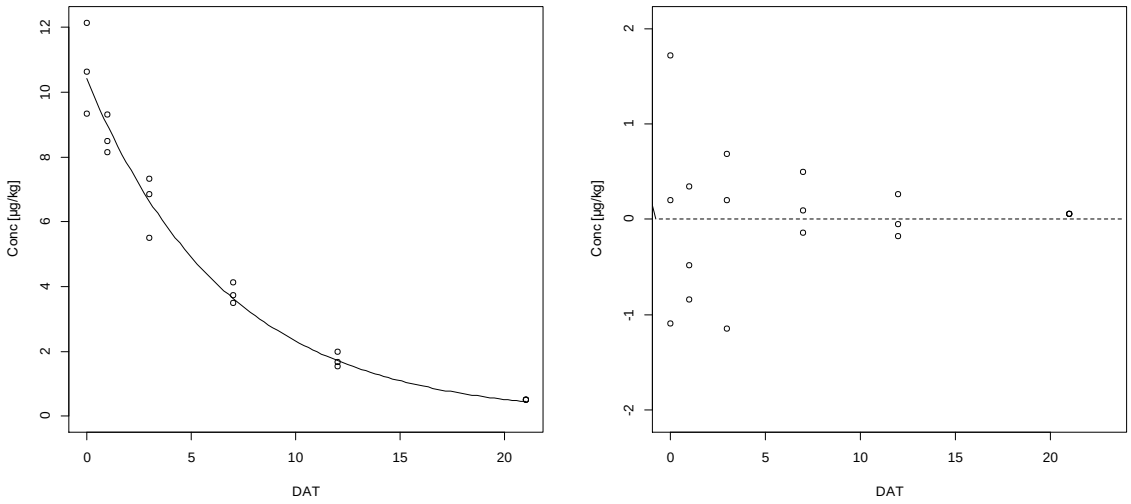



**Figure F2**: Dissipation curve fitting (left) and related residuals (right)

Model parameters (SFO)

| **Parameter** | **Mean estimate** | **95 % Confidence interval** | ***p*-value (*t*-test)** |
|---|---|---|---|
| M0 (initial value) [µg/kg] | 10.43 | 9.78–11.1 | <0.0001 |
| K (dissipation constant) [day$^{-1}$] | 0.15 | 0.13–0.18 | <0.0001 |

Model outcome

| | |
|---|---|
| $DT_{50}$ [days] | 4.6 |
| $DT_{90}$ [days] | 15.3 |
| $\chi^2$ % | 2.84 |

*Evaluation of the fitting*

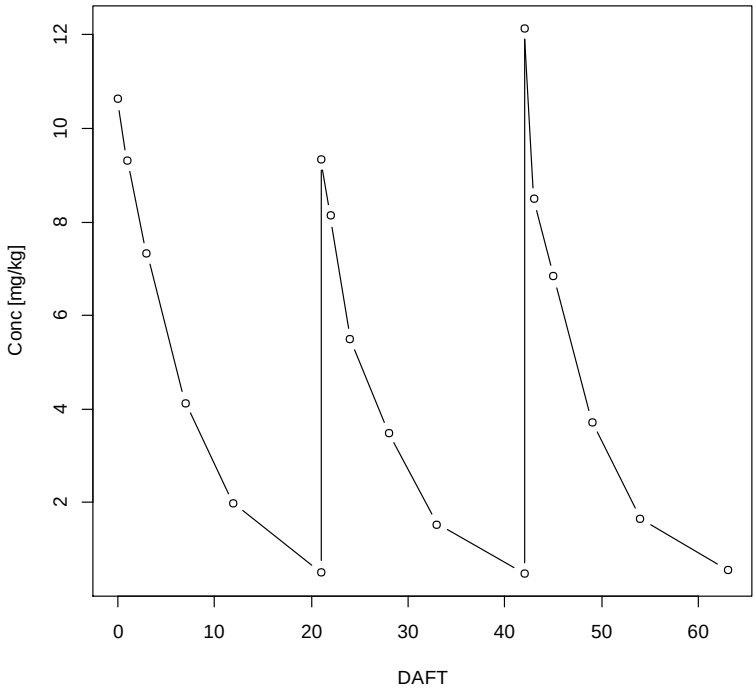
1- Visual fit → plot of time vs concentration shows that the fitted line should pass in the vicinity of the measurement points for all dates.
2- Residual plot → Plot of time vs residuals against the $y = 0$ line shows that points are scattered around the zero line. No regular patterns are identified and no systematic underestimation or overestimation is present at any date.
3- Chi-square % → well below 15 %.
4- *t*-test for individual model parameters → *t*-test resulted in *p*-values well below 0.05, indicating high confidence in the estimation of the model parameters.

## Worked example B: Dealing with multiple applications

Within this worked example two different options are presented to derive a $DT_{50}$ from a single residue trial with multiple applications. The available fake dataset is supposed to contain residue data for a random pesticide applied three times to pome fruits, with an interval of 21 days. Results are related to fruit.

| DAFT | Concentration [mg/kg] |
|---|---|
| 0 | 10.64 |
| 1 | 9.32 |
| 3 | 7.33 |
| 7 | 4.14 |
| 12 | 1.98 |
| 21 | 0.51 |
| 21 | 9.35 |
| 22 | 8.14 |
| 24 | 5.50 |
| 28 | 3.50 |
| 33 | 1.54 |
| 42 | 0.48 |
| 42 | 12.15 |
| 43 | 8.50 |
| 45 | 6.84 |
| 49 | 3.73 |
| 54 | 1.67 |
| 63 | 0.56 |

DAFT: Days after first application




*Option 1: consider each application as a standalone trial*

The first procedure presented in Section 3 is followed in the example below.

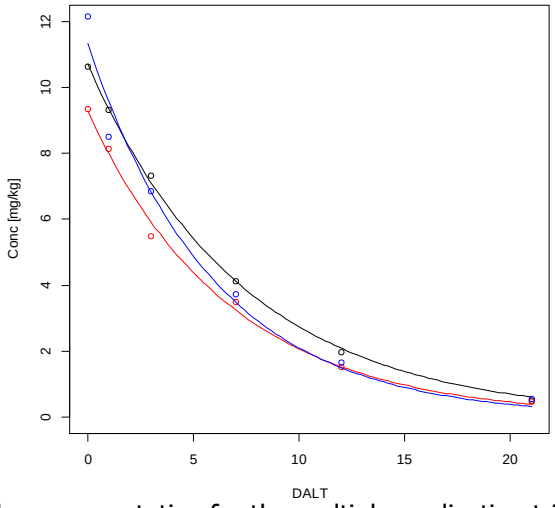
1- Consider each application (and the following points until the next application) as a standalone trial.

| DALT | Concentration [mg/kg] | | |
|---|---|---|---|
| | Sub-trial 1 | Sub-trial 2 | Sub-trial 3 |
| 0 | 10.64 | 9.35 | 12.15 |
| 1 | 9.32 | 8.14 | 8.50 |
| 3 | 7.33 | 5.50 | 6.84 |
| 7 | 4.14 | 3.50 | 3.73 |
| 12 | 1.98 | 1.54 | 1.67 |
| 21 | 0.51 | 0.48 | 0.56 |

DALT= Days after last application

2- Calculate as many $DT_{50}$ as the number of applications.

| Sub-trial | $DT_{50}$ [days] |
|---|---|
| Sub-trial 1 (black line) | 5.09 |
| Sub-trial 2 (red line) | 4.62 |
| Sub-trial 3 (blue line) | 4.11 |




Calculate the geomean of the calculated $DT_{50}$ as the representative for the multiple application trial.

| Sub-trial | $DT_{50}$ [days] |
|---|---|
| Sub-trial 1 (black line) | 5.09 |
| Sub-trial 2 (red line) | 4.62 |
| Sub-trial 3 (blue line) | 4.11 |
| Geomean | **4.59** |

*Option 2: consider all applications in a unique fitting*

The second procedure presented in Section 3 is followed in the example below.

1- Express all concentrations in terms of fraction of the one measured on the last application date (i.e. on the day of each application, the value will be 1).

| DAFT | Concentration [mg/kg] | Last application on DAFT | Concentration as fraction of 0 DALT [adimensional] |
|---|---|---|---|
| 0 | 10.64 | 0 | 1.00 |
| 1 | 9.32 | | 0.88 |
| 3 | 7.33 | | 0.69 |
| 7 | 4.14 | | 0.39 |
| 12 | 1.98 | | 0.19 |
| 21 | 0.51 | | 0.05 |
| 21 | 9.35 | 21 | 1.00 |
| 22 | 8.14 | | 0.87 |
| 24 | 5.50 | | 0.59 |
| 28 | 3.50 | | 0.37 |
| 33 | 1.54 | | 0.16 |
| 42 | 0.48 | | 0.05 |
| 42 | 12.15 | 42 | 1.00 |
| 43 | 8.50 | | 0.70 |
| 45 | 6.84 | | 0.56 |
| 49 | 3.73 | | 0.31 |
| 54 | 1.67 | | 0.14 |
| 63 | 0.56 | | 0.05 |

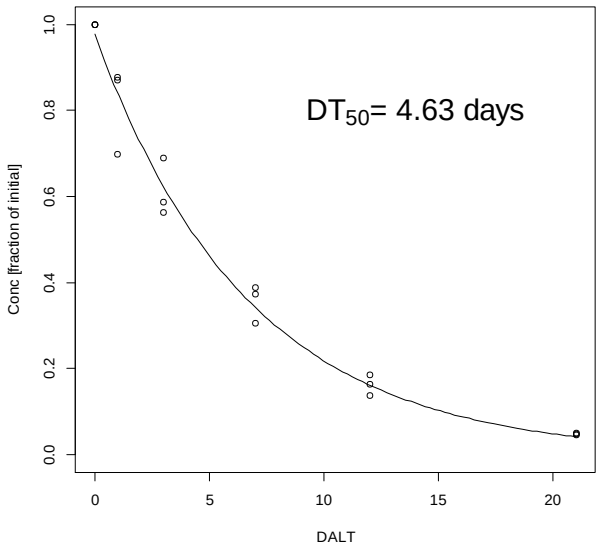
DAFT= Days after first application; DALT= Days after last application

2- Calculate the time between each measurement time point and the date of the last application.

| DALT | Concentration as fraction of 0 DALT [adimensional] |
|---|---|
| 0 | 1.00 |
| 1 | 0.88 |
| 3 | 0.69 |
| 7 | 0.39 |
| 12 | 0.19 |
| 21 | 0.05 |
| 0 | 1.00 |
| 1 | 0.87 |
| 3 | 0.59 |
| 7 | 0.37 |
| 12 | 0.16 |
| 21 | 0.05 |
| 0 | 1.00 |
| 1 | 0.70 |
| 3 | 0.56 |
| 7 | 0.31 |
| 12 | 0.14 |
| 21 | 0.05 |

DALT= Days after last application

3- Use the newly derived values for the fitting exercise.



## Worked example C: Dealing with toxic metabolites

If the toxicity of the metabolite is comparable to that of the parent, two approaches can be followed.

**Simpler approach**

The simpler approach can be followed when it can be assumed that:

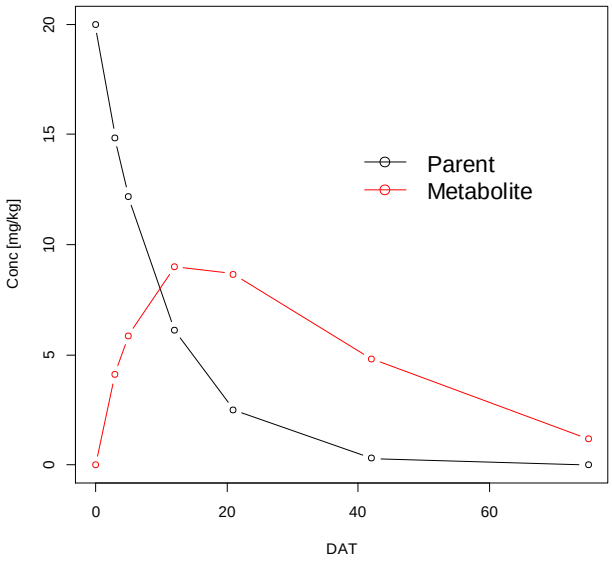
*toxicity of the parent ≈ toxicity of the metabolite (within a factor 2–3)*

In this case it is enough to sum the residue concentrations at any time point of the parent and the metabolite. Then fit the obtained data as if it was a single chemical.

Let's assume we have a dataset reporting measured residues of the parent and one metabolite of equal toxicity. The measurements were performed on six dates, from DAT 0 to DAT 75.

| DAT | Concentration [mg/kg] | |
|---|---|---|
| | Parent | Metabolite |
| 0 | 20 | 0 |
| 3 | 14.86 | 4.11 |
| 5 | 12.19 | 5.86 |
| 12 | 6.1 | 9.02 |
| 21 | 2.5 | 8.66 |
| 42 | 0.31 | 4.81 |
| 75 | 0.01 | 1.18 |

DAT: days after treatment.

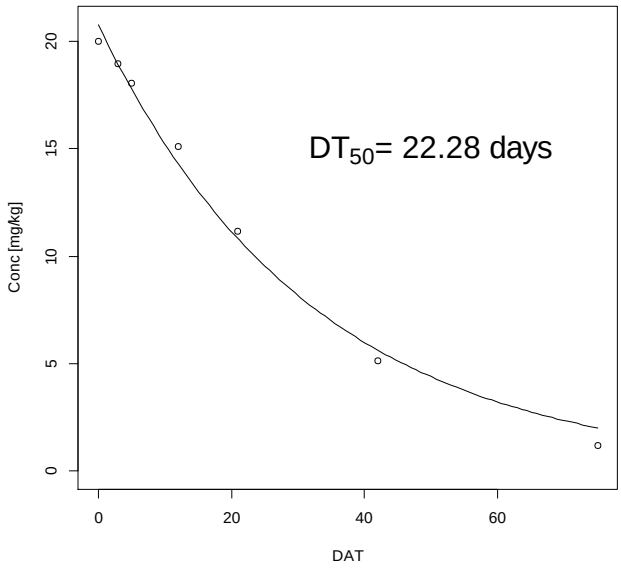



Since the toxicity of the two compounds is practically equal, it makes sense to sum the concentration at each time point.

| **DAT** | **Concentration [mg/kg]** | | |
|---|---|---|---|
| | **Parent** | **Metabolite** | **Sum** |
| 0 | 20 | 0 | 20 |
| 3 | 14.86 | 4.11 | 18.97 |
| 5 | 12.19 | 5.86 | 18.05 |
| 12 | 6.1 | 9.02 | 15.12 |
| 21 | 2.5 | 8.66 | 11.16 |
| 42 | 0.31 | 4.81 | 5.12 |
| 75 | 0.01 | 1.18 | 1.19 |

DAT: days after treatment.

Once calculated, the sum can be used for fitting a decline curve.




Then calculate the combined $f_{TWA}$ using the combined rate constant:
$f_{TWA} = (1-e^{-kt}) / kt$
where $k = \ln(2) / DT_{50}$
Assuming $t$=21 d: $f_{TWA} = 0.73$

**More complex approach (parent-normalised $f_{TWA}$)**
This approach should be followed when the toxicity of the metabolite and that of the parent are comparable (less than a factor of 10), but cannot be assumed to be equivalent. It is anticipated that the following approach will only be used when:

- the metabolite has a high formation fraction and it is more persistent than the parent
- the metabolite is formed in a medium-high amount and is more toxic than the parent.

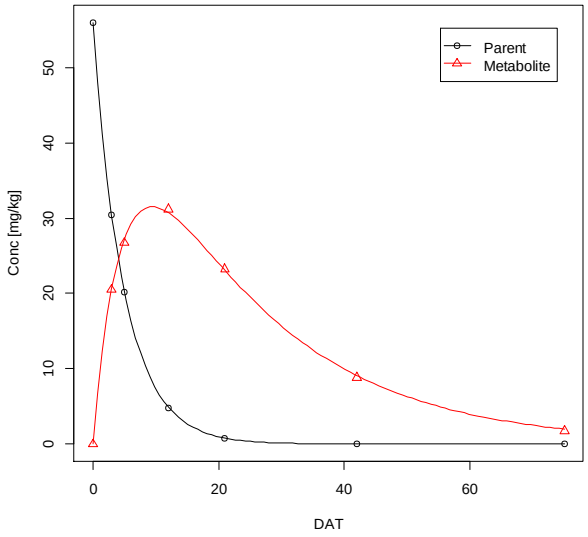
The example dataset includes measured residues of the parent and one metabolite. The measurements were performed on six dates, from DAT 0 to DAT 75. The toxicity of the metabolite is known to be equivalent to one quarter of that of the parent.

| DAT | Concentration [mg/kg] Parent | Metabolite |
|---|---|---|
| 0 | 56 | 0 |
| 3 | 30.38 | 20.5 |
| 5 | 20.21 | 26.7 |
| 12 | 4.85 | 31.17 |
| 21 | 0.77 | 23.23 |
| 42 | 0.01 | 8.82 |
| 75 | 0 | 1.73 |

DAT: days after treatment.

1- Estimate the kinetic parameters for the parent and (optionally) for the metabolite (in which case the kinetics fit should be done for the two substances together).

| Parameter | Parent | Metabolite |
|---|---|---|
| Kinetics model | SFO | SFO |
| M0 (concentration at *t*=0) [mg/kg] | 55.89 | - |
| $DT_{50}$ [days] | 3.42 | 15.02 |
| Formation fraction | - | 0.87 |
| $\chi^2$ % | 0.44 | 1.22 |




2- Calculate the $f_{TWA}$ of the parent and of the metabolite between DAT 0 and DAT 21. Care should be taken as the $f_{TWA}$ of the metabolite cannot be derived analytically with the standard equation $(1-e^{-k*Time})/(k*Time)$. On the contrary, this should be quantified as AUC divided by the length of the time window. The $f_{TWA}$ for the metabolite is calculated as the TWA concentration of the metabolite (same interval used for the parent) divided by the initial concentration of the parent.

$f_{TWA}$ Parent = 0.23
$f_{TWA}$ Metabolite = 0.41

3- Recalculate the $f_{TWA}$ of the metabolite accounting for the toxicity ratio with the parent (in this case the ratio = 0.25) and finally sum up this value with the $f_{TWA}$ for the parent.

Total $f_{TWA}$ = (0.41 * 0.25) + 0.23 = 0.347

# Appendix G – How to express the endpoint for sediment-dwelling organisms

## Example of mass balance calculations

The starting point for a mass balance calculation for a chronic study with sediment-dwellers is the measured concentrations of the active substance in the test system encompassing the concentration at the beginning and end of the study.

The information on the analytical determination of the active substance needed for the mass balance calculation in the case of a sediment-dweller study (spiked water) is exemplified in Table G1. It is noted that for simplicity the measurements in this case were reported at two time points only (beginning and end of the study). However, measurements at additional time points are recommended, particularly in the case of the substances that are difficult to test (concentrations are poorly maintained in the test system).

**Table G1:** Measured concentrations of the active substance in a spiked water study with *Chironomus riparius* performed in line with OECD TG 219 (OECD, 2004b)

| Measured concentrations | | | | | | |
|---|---|---|---|---|---|---|
| | 1 hour | | | 28 days | | |
| **Nominal (µg a.s./L)*** | **Overlying water (µg a.s./L)** | **Pore water (µg a.s./L)** | **Sediment (µg a.s./kg)** | **Overlying water (µg a.s./L)** | **Pore water (µg a.s./L)** | **Sediment (µg a.s./kg)** |
| Control | <LOQ | <LOQ | <LOQ | <LOQ | <LOQ | <LOQ |
| Solvent control | <LOQ | <LOQ | <LOQ | <LOQ | <LOQ | <LOQ |
| 1.5 | 0.99 | 0.077 | 1.5 | 0.05 | 0.03 | 1.3 |
| 5 | 3.5 | 0.16 | 3.6 | 0.1 | 0.053 | 2.7 |
| 10 | 6.8 | 0.35 | 8.2 | 1.5 | 0.067 | 7.3 |

a.s.: active substance; LOQ: limit of quantification
* spiked water study.

Starting from the information in Table G1 and considering the amount of sediment (0.1 kg) and volume of water (300 mL) used in the test system, it is possible to calculate the amount of active substance that partitions to the various compartments by multiplying the concentration by the volume/amount of water/sediment and to calculate the percentage in each compartment with respect to that initially applied. In the example, in the absence of the volume of the pore water, the amount in pore and overlying water were considered together (see Table G2).

In order to assess the fate of the active substance in the sediment/water system the mass balance calculation was performed for each time point for which measurements were available (see Table G2).

**Table G2**: Example of mass balance calculations for an active substance. The amount in each compartment is expressed in terms of percentage of active substance with respect to the initially applied amount in water.

| Mass balance* | | | | | | | |
|---|---|---|---|---|---|---|---|
| | 1 hour | | | 28 days | | | |
| **Nominal (µg a.s.)** | **Overlying water (%)** | **Pore water (%)**** | **Sediment (%)** | **Overlying water (%)** | **Pore water (%)**** | **Sediment (%)** | **Total** |
| Control | / | / | / | / | / | / | / |
| Solvent control | / | / | / | / | / | / | / |
| 0.45 | 71 | / | 33 | 5% | / | 29% | 34% |
| 1.5 | 73 | / | 24 | 3% | / | 18% | 21% |
| 3 | 72 | / | 27 | 16% | / | 24% | 40% |

a.s.: active substance

* Considering that the amount of sediment included in the text system was 0.1 kg and the volume of water 300 mL. The volume of pore water was unknown, as is frequently the case.

** The volume of pore water was not available, therefore the amount in pore and overlying water were considered together; this creates an uncertainty in the calculations.

In the example, the active substance partition to the sediment is up to 33 % in the first hour, while at the end of the study the substance is still present in the sediment at levels of up to 29 % of that initially applied. In the case of the water compartment, both dissipation to the sediment and degradation occurred; the total amount of the active substance in the test system in the study accounts for up to 40 % of the applied amount.

In this case it is recommended that the endpoints are expressed in terms of the geometric mean of measured concentrations considering that the concentrations were not maintained and to consider the dissipation in the sediment. In order to better calculate geometric mean measures, intermediate analytical measurements are highly recommended. It is additionally recommended that the key endpoints are presented in terms of mg of active substance per kg of dry sediment and mg of active substance per L of water. This would ensure that both exposure via water and sediment are covered for sediment dwellers.

A similar approach to the one reported above could also be used for studies with spiked sediment.

# Appendix H – Use of de Jong et al. (2010) for non-target arthropod field studies

Field tests with NTAs are carried out in accordance with ESCORT 2 and Candolfi et al. (2000a, 2000b) that discuss the experimental conditions, treatment, application and sampling, data analysis and reporting. However, details and criteria on how to evaluate field studies were missing. In 2010, guidance on how to summarise those studies was published by de Jong et al. in order to provide a basis for a detailed and harmonised evaluation of non-target field studies.

The guidance gives recommendations on a number of items which should be considered when assessing the reliability of a field study with NTAs. Three reliability indices are proposed ('reliable', 'less reliable' and 'not reliable'). Ten test items to be evaluated are listed in the table below, including: information on the substance, the test site, the application, the experimental test design, the biological system and the sampling. Under results, items related to application, endpoint and elaboration of results are proposed. For each item, a set of questions is proposed which can guide the assessment of reliability.

**Table H1:** For each item proposed for the evaluation, recommendations are provided on issues to consider when evaluating different relevant items of field studies with arthropods

| No | Test item | Recommendations |
|---|---|---|
| 1 | Substance (formulation, toxic reference, etc.) | Information about the applied substance (active substance or formulation) and the toxic reference (if used) should be reported. The guidance specifies that the same test item can be a reference when used at a higher application rate (able to cause 50 % effects): Clear effects should be found in the toxic reference, at least a 50 % effect on at least one sampling date, for at least 10 % of the taxa for which statistical evaluation is possible, and when these criteria are not met the test is not reliable.<br><br>When no reference item is included, the highest application rate of the test item could act as such, and in that case the same criteria are used for the highest treatment rate as for the reference item.<br><br>In the case that a toxic reference item is not included, high enough rates of the test item should be applied to cause clear effects as a toxic reference, unless effects were clearly seen with the test item at the 'target' application rate(s). If not, the study should be classified as 'unreliable'. This is in agreement with what is written in the guidance. It should be noted that the test with a toxic reference item is a validation tool.<br><br>The use class (e.g. insecticide, herbicide) and mode of action (e.g. contact, systemic, cholinesterase inhibitor) of the test item should be reported. |
| 2 | Test site | The history of the test site at least two years before the start of the experiment should be available (e.g. previous cropping history, application of pesticides, mineral fertilisers, establishment of orchards, crop rotation for arable crops, etc.).<br><br>Treatments applied to maintain the health of the crop, e.g. fungicides, must be applied to the whole test site. When the results of a field study should be used for assessment of the potential impact on the off-crop fauna, the off-crop area is considered to be an undisturbed area (use of other pesticides is not acceptable). |
| 3 | Application | Data about the application are relevant in order to evaluate whether the application in terms of mode of application, dosage, number of applications and interval between applications, reflects the GAP. Information on the climatic conditions in the period before, during and after the application as well as information about artificial irrigation should also be reported. The field study should preferably be conducted in the season of the proposed use of the substance. The above are important to evaluate the correct exposure of NTAs to the tested substance. |
| 4 | Experimental test design | Random plot design, Latin square, plot size (a minimum plot size of 1 ha for arable land and 0.2 ha for orchards is recommended), number of replicates, number of samples.<br><br>Recovery could differ for off-crop and in-crop sites. In terms of recovery, the scale of the study should be considered when comparing it with the scale of the field under the proposed use. The duration of the study should be long enough in order to assess the recovery within the test period. Recovery is assessed for different taxonomic levels, from population to community. Delayed effects may occur after recovery has been demonstrated and after the test period. This is relevant for sensitive life stages and should be addressed in the study report. It should be noted that regarding the issue of whether the potential for recovery/recolonisation should be demonstrated to be below one year, more criteria are needed. |

| | | |
|---|---|---|
| | | An increasing number of field studies are conducted under the principles of good laboratory practice. For new studies this is a requirement. |
| 5 | Biological system | For the time being, a quite extensive and detailed list of taxa is provided in Table 4 of the de Jong et al. (2010) guidance for reliability assessment and is agreed as a minimum requirement for arable crops, orchards and off-field. Thus an updated version of Table 4, including the previously missing footnotes, should be used as a reference for the reliability assessment, as is included in the meeting technical report. In agreement with the de Jong et al. (2010) guidance, the desired taxa level of identification is provided in this table; about 50–80 taxa are available to allow for statistical analysis with sufficient power in a typical field study. Also, the minimum number of individuals should fulfil the requirements of statistical analysis.<br><br>It should be noted that if the listed taxa are lacking, a study is not invalidated, hence the evaluator should clarify the issue; e.g. seek a justification for the lacking/not measured or additionally reported taxa under local conditions. The biological system should be summarised (e.g. dominant groups, the frequency of species found, etc.). |
| 6 | Sampling | Sampling method, scheme, area, etc. Some general guidance is given in Candolfi et al. (2000a). In the study report it should be clearly indicated which sampling method is used for each group of species.<br><br>Given the (sometimes) large variability of a population over time, the pre-treatment monitoring of the community should be conducted not too long before treatment. Pre-treatment sampling, preferably shortly (< 5 days) before the first application, is desired in order to assess the variation between plots and the taxa exposed. In some cases (e.g. application early in the growing season or in the winter) this is not useful or possible because certain organisms are not yet present in sufficient numbers.<br><br>Weather conditions in the period before sampling should be recorded.<br><br>For off-crop risk assessment the populations of organisms living on the soil surface should be recorded as well. |
| 7 | Results in terms of application | According to the guidance, it should be possible to check whether the right amount of the substance studied was applied in the test: e.g. by measuring the compound in the spray solution and controls of the spray pattern. The weather conditions during the test should be considered, and attention should be paid to deviations from the average conditions of the test site (e.g. heavy rainfall or unusually low or high temperatures on the day of application that could influence exposure of the NTA fauna). |
| 8 | Endpoints | Population level effects should be reported. The population effect on each taxon including sensitive life stages and, where possible, recovery with time to recovery, compared to controls should be reported.<br><br>Number of arthropod species/taxa and individuals and community groups (e.g. Aranae, Insecta, etc.; juveniles and adults, separately).<br><br>Total biomass of all arthropod and community groups (e.g. Aranae, Insecta, etc.; juveniles and adults, separately).<br><br>Numbers and biomass of at least the two most abundant species/taxa (juveniles and adults, separately).<br><br>Functional endpoints: e.g. parasitism rates.<br><br>Indirect effects: e.g. prey items counted to interpret the importance of food/prey removal.<br><br>Depending on the test design, an assessment endpoint could be derived (no |

| | | |
|---|---|---|
| | | observed effects rate, no observed ecological adverse effects rate, lowest observed ecological adverse effect rate). |
| 9 | Elaboration of the results | Statistical analysis<br>Multivariate or univariate (ANOVA) techniques can be used.<br>It is recommended that a power analysis is always provided for the endpoints investigated in the study.<br>When elaborating the results, consideration should be given to biological relevance vs statistical significance of observed effects.<br>The concept of MDD refers to the magnitude of the effect that needs to exist in the treatment population in relation to the control in order to obtain a statistically significant difference in hypothesis testing. The MDD concept is potentially very beneficial for the interpretation of the field studies, but further criteria need to be developed specifically for NTAs in order to fruitfully use this information in the assessment.<br>Community analysis tools such as principal response curve could be used but should not be specifically requested (optional).<br>Summary Table 2 in de Jong et al. (2010) is useful for a quick overview of effects and should be included. However, more details in a less aggregated form have been provided in the study summary in order to allow for a transparent evaluation. |
| 10 | Effect classification | For the effects, a classification is recommended on page 25, Table 5 of de Jong et al. (2010). However, the effect classes are not considered for the time being. It is optional to report them but if they are missing from the report it would not lead to a lowering of the reliability score. The proposal of using effect classes can be further considered in future development activities. (e.g. EFSA PPR Panel, 2015). |

### Footnotes and legend to Table 4 on page 22 of the guidance by de Jong et al. (2010)

The footnotes and legend for Table 4 on p. 22 are missing from the de Jong et al. (2010) guidance document. The information was received from the authors of the guidance. In order to facilitate the use of Table 4 of the guidance it is provided below:

** For Coccinellidae the remark has to be made that species from this taxon can populate a certain area relatively quickly as a result of the presence of aphids. When aphids are not present and abundant, Coccinellidae will not appear; this does not render the test directly unreliable, however this phenomenon should be taken into account when evaluating the study.*

*Legend:*

*'+' means that the taxon should be present and identified at the level specified, else the test is not sufficiently comprehensive to be of general validity. When '+' taxa are lacking in the specified agro-ecosystem addition of appropriate data, for example from other (laboratory) studies is needed to make the test reliable, otherwise the test is considered unreliable.*

*A '+/-' means that a taxon should be present in the south of Europe, but not necessarily in the north of Europe.*

*A '0' means that the test is less reliable (Ri 2) when sufficiently robust data at the indicated level of taxonomic precision are missing, but additional data are not required.*

*A '-' indicates that a specified taxon is generally not relevant for the specified cropping system(s).*

*'Off-crop' means non-cropped lands in the vicinity of agricultural fields, e.g. meadows or woodlands.*

### Additional points from the meeting discussion

- **Summarising the results of an arthropod field study**

Summary Table 2 (p. 17 of the de Jong et al. (2010) guidance) should always be included in the study summary at the level of aggregation/detail that is proposed in the de Jong et al. (2010) guidance. Also, less aggregated data, significance levels of effects and the size of the detected effect would be suitable to be demonstrated in the table. Experts agreed that summary Table 2 should be included in the NTA study summaries at the level of aggregation/detail that is proposed in the de Jong et al. (2010) guidance.

- **List of taxa for reliability assessment**

The list of taxa for reliability assessment should be used as a reference for arable crops, orchards and off-field (item 5, Table 4, p. 22 of the guidance).

- **Toxic reference item**

In agreement with the guidance, the applicant is to include a toxic reference item or to apply rates of the test item high enough to cause clear effects as a toxic reference, unless effects were clearly seen with the test item at the 'target' application rate(s). If not, the study should be classified as 'unreliable'.

- **MDD concept and criteria**

Further criteria of the MDD concept need to be developed specifically for NTAs in order to fruitfully use this information in the assessment. It is recommended that a power analysis is provided together with the study by the applicant.

- **Principal response curves**

The principal response curves can be optionally (but not specifically) provided.

- **Effect classes**

The effect classes can be provided but are not required. This can be further considered in future development activities.

- **Recovery**

Overall, the meeting raised the concern that aged residue studies as such might not be sufficient to demonstrate recovery or recolonisation as recolonisation is highly dependent on the landscape configuration and the species' traits. Regarding the maximum time considered acceptable for ageing of residues, it was concluded that more criteria are needed to demonstrate whether the potential for recovery/recolonisation should be below one year.

# Appendix I – Use of de Jong et al. (2006) guidance for soil organisms

Earthworm field tests are carried out according to ISO 11268-3 (2014). In 2006, guidance on how to summarise those studies was published by de Jong et al. (2006).

The guidance gives recommendations on several items which should be considered when assessing the reliability of an earthworm field study. Three reliability indices are proposed ('reliable', 'less reliable' and 'not reliable') which could also be extended to four ('reliable', 'reliable with minor restrictions', 'reliable with major restrictions' and 'not reliable'). The test items to be evaluated are categorised in two big macro-categories: description and results. Within the description macro-category six items are listed: information on the substance, the test site, the application, the experimental test design, the biological system and the sampling. Under results, items related to application, endpoint and elaboration of results are proposed. For each item, a set of questions which may guide in the assessment of reliability is proposed.

A draft OECD guideline which is based on the current ISO guideline, is under development. The OECD guidelines will provide more information on aspects such as NOEC and ECx designs, exposure assessment and statistical evaluation of the results.

The agreed approaches on how to assess the reliability of earthworm field tests might, therefore, need to be adjusted once the new OECD guideline comes into force.

**Table I1:** For each item proposed for the evaluation of earthworm field studies according to de Jong et al. (2006), recommendations are provided which are in line with the current ISO 11268-3 (2014) and Römbke et al. (2006).

| No | Test item | Recommendations |
|---|---|---|
| 1 | Substance (formulation, vehicle, reference item, etc.) | Information about the applied substance and the toxic reference should be reported. For the reference substance, an application of 6 to 10 kg carbendazim is considered appropriate. |
| 2 | Test site | The history of the test site should be available. According to the ISO guideline the description of the test site should include:<br>— particle-size distribution (as specified in ISO 11277 (ISO, 2009))<br>— organic-carbon content (as specified in ISO 10694 (ISO, 1995))<br>— pH-value (as specified in ISO 10390 (ISO, 2005))<br>— water-holding capacity, WHCmax (in the A-horizon, as specified in ISO 11274 (ISO, 1998))<br>— description of vegetation<br>— history of the test site (e.g. application of PPPs in the previous years, particularly PPPs with similar modes of action).<br>Moisture content is one of the only parameters which is considered to change considerably over the course of the study and therefore it is suggested to monitor it in parallel to the biological sampling.<br>Climatic conditions such as temperature and rainfall (monthly figures could be suitable) should also be reported.<br>Grassland or arable field can be used. However, grassland should be the preferred study site for testing the effects of substances on earthworms. In grassland, earthworm density and diversity are generally higher and more stable than on arable land. However, to overcome the issue of the crop interception that may differ between grassland and arable field, it is suggested that grass should be cut before application. The disturbance of the site should be kept to a minimum. Orchards are not recommended for testing because of the heterogeneity of the site due to tree rows and strips without trees. |
| 3 | Application | Data about the application are relevant in order to evaluate whether the application in terms of mode of application, dosage, number of applications and interval between applications, reflects the GAP. If an accumulation of the tested substance in the soil is modelled, the PEC background should also be considered in the study (e.g. through incorporation in the soil half a year before study start).<br>Information on the climatic conditions in the period before, during and after the application as well as information about irrigation should also be reported. This is important to evaluate the correct exposure of earthworms to the tested substance. |
| 4 | Experimental test design | Random plot design, plots of at least 100 $m^2$, with a treated 1–2 m edge strip, four replicate plots at least for each test variant (i.e. control, treatment and reference item) and at least four subsamples per plot.<br>The envisaged statistical power of the test (see below under '9. Elaboration |

| | | |
|---|---|---|
| | | of the results') should be considered in the study set-up (e.g. number of plots and/or samples per plot).<br><br>The duration should be one year for assessing recovery. However, if a compound is applied in autumn it is recommended to prolong until the next cropping season. The application time for products should be in line with GAP. These recommendations, however, may change in future, once the risk assessment for soil organisms has been be revised (e.g. need for testing more than one application rate for the determination of an endpoint). |
| 5 | Biological system | The test area should present an earthworm density of at least 60 ind/$m^2$ for arable sites and 100 ind/$m^2$ for grassland. A mixed community should be present; *Lumbricus* spp. and *Aporrectodea caliginosa* are considered the typical dominant species in agricultural areas. In some areas, however, *A. caliginosa* is not dominant, but e.g. *Allolobophora chlorotica* is the dominant one. Therefore, it is important that at least two ecological groups (i.e. anecic, endogeic, epigeic) are present with at least one species having 10 % dominance. In this respect, information from the studies by, e.g., Dinter et al. (2013) and Van Capelle et al. (2016) that describe the occurrence and distribution of earthworms in agricultural landscapes across Europe could be useful. |
| 6 | Sampling | On grassland, a sampling area of 0.25 $m^2$ per individual sample is currently considered sufficient; while on arable land, the sample area should usually be increased to 1 $m^2$ due to low population density or non-homogeneous distribution of the worms. In grassland the vegetation at the sampling area should be cut before sampling; sampling should be at least taken 1, 4, 6 and 12 months after the application. The time of the sampling should include the active peaks of earthworms in spring (April/May) and autumn (September/October). Pre-treatment sampling should not be done too long before the treatment (e.g. two weeks before substance application), due to the high temporal variability. Pre-sampling should be performed after the last management of the area, e.g. after mowing, in order to determine the correct abundance at test start.<br><br>For sampling earthworms, the formaldehyde extraction method, the mustard method or the octet method have been used. In ISO 11268-3 (2014) (version of 2014), reference is made to ISO 23611-1 (ISO, 2006) regarding the earthworm extraction methods. The guideline has, however, been reviewed lately (2018). In ISO 23611-1, version 2011, a combination of hand-sorting and formalin extraction was recommended. However, the updated ISO 23611-1 (2018) replaces the formalin extraction with an extraction employing AITC (allyl-isothiocynate, the active substance of mustard). Formalin extraction is no longer recommended.<br><br>Moreover, the octet extraction method is also considered outdated, considered to be inefficient. This method inappropriately reflects the actual community structure (poor extraction of anecic species) under dry conditions and the efficiency was not improved by water addition beforehand (Eisenhauer et al., 2008). Furthermore, another problem for the efficiency of the octet method may be the inhomogeneous soil structure (Čoja et al., 2008).<br><br>The adult and juvenile worms should be counted separately. Adults should be identified at the species level while juveniles should be distinguished between tanylobous and epilobous species. Earthworms should also be classified as anecic, endogeic and epigeic (if any). |
| 7 | Results in terms of application | Immediately after application, the concentration of the test substance in soil should be determined once by residue analysis to verify the actual exposure |

| | | |
|---|---|---|
| | | concentration in soil. For soil sampling, the (OECD, 2016) can be followed.<br><br>Soil cores should be cut into different layers (0–1; 1–3*; 3*–5; 5–10 and 10–20 cm or into 0–5; 5–10 and 10–20 cm segments).<br><br>Considering the wide variability in field studies, a range of 50 to 150 % of the nominal concentration in soil should be achieved for quality assurance measures. Possibly, the verification of the test concentration should be carried out not only for verifying the application rate (as recommended by the guideline) but also for comparing the measured concentration with the predicted one. This would allow an assessment of whether earthworms were correctly exposed.<br><br>When available, it is recommended that the measured residues are expressed in quantities comparable with the PEC, e.g. over 1, 2.5 or 5 cm. Check whether the initial measured concentrations cover the PEC calculated for the intended uses.<br><br>It would be desirable to sample soil for residue analysis in parallel to the earthworm samplings. However, this is not required by the ISO guideline. |
| 8 | Endpoints | Total abundance of earthworms and tanylobous/epilobous individuals (juveniles and adults, separately).<br><br>Total biomass of all earthworms and biomass of tanylobous and epilobous individuals (juveniles and adults, separately).<br><br>Abundance of the determined species (adults including at least the dominant species).<br><br>Biomass of the determined species (adults).<br><br>Species diversity as taxa richness. |
| 9 | Elaboration of the results | **Statistical analysis**<br><br>In order to test for normality and variance homogeneity, Shapiro–Wilks and Levene's test procedures are recommended to be used, respectively. With normally distributed and homogeneous data, Dunnett's or Williams' test ( $\alpha$ = 0.05, one-sided) should be performed. If data do not fulfil the criterion of normality, they can be transformed (logarithmic, square-root) or evaluated using generalised linear models or non-parametric tests, e.g. the Bonferroni U-test or the Jonckheere–Terpstra step-down test can be applied. If only one treatment has been performed and the prerequisites (normality, homogeneity) of the parametric test procedures are fulfilled, the pairwise Student's *t*-test, or otherwise the Mann–Whitney U-test procedure can be used.<br><br>It should be noted that data from different treatments often do not fulfil the requirements of variance homogeneity (e.g. following very strong effects, the variance of a treatment will be 0). Therefore, parametric tests cannot be used. The use of non-parametric tests (e.g. U-test) often implies a lower discrimination power. New approaches are currently discussed and will be included in the upcoming OECD guideline.<br><br>Regarding the statistical power of earthworm field studies, little guidance is available on how to estimate it.<br><br>The minimum detectable difference (MDD) as used for the evaluation of the micro/mesocosm experiments could be extended to terrestrial higher tier studies. The concept of MDD refers to the magnitude of the effect that needs to exist in the treatment population in relation to the control in order to obtain a statistically significant difference in hypothesis testing. |

| | | |
|---|---|---|
| | | Five classes are defined for the interpretation of the MDD results.<br><br>**Class** / **MDD** / **Comment**<br>0 / > 100 % / No effects can be determined<br>I / 90–100 % / Only large effects can be determined<br>II / 70–90 % / Large to medium effects can be determined<br>III / 50–70 % / Medium effects can be determined<br>IV / < 50 % / Small effects can be determined<br><br>An attempt to estimate the MDD of an earthworm field study by using the MDD concept (EFSA PPR Panel, 2013; Brock et al., 2015) was presented at the SETAC conference 2018 (Bayona et al., 2018). A paper on the same topic has been published (Andrade et al., 2017). The conclusion of the two available papers is not consistent. Andrade et al. (2017) concluded that small effects on overall earthworm abundance and biomass can be consistently detected with a good degree of statistical confidence and small to medium effects are often also detectable in the case of species-specific variables, while Bayona et al. (2018) doubted the robustness of the assessed study since no effects were observed at community level. In addition, the statistical power of the test was not considered sufficient to detect effects as the MDD was higher than 100 % in 50 % of sampling dates.<br><br>de Jong et al. (2006) also addressed the limitations of effect detection < 50 % in earthworm field studies due to the high natural variability. In order to increase the statistical power of earthworm field studies, the set-up in terms of numbers per plot or subsamples per plot might be further improved in the upcoming OECD guideline.<br><br>**Community endpoint evaluation**<br><br>In addition to the above, a community analysis tool such as the principal response curve could be used.<br><br>Species diversity analyses (e.g. the Shannon–Wiener index to describe the taxa richness as well as frequency distribution) as well as similarity analysis (e.g. the Steinhaus index to describe the similarity of communities between different treatments) might help in the interpretation of the results.<br><br>Performing community endpoint evaluations will also include available results from those species with low abundance and/or steadiness that cannot be addressed in univariate analyses. |
| 10 | Biological relevance versus statistical significance | As described above (point 9), the detection of statistically significant effects in earthworm field studies is often hampered by the low statistical power of the tests.<br><br>EFSA Guidance on the assessment of the biological relevance of data in scientific assessments (EFSA Scientific Committee, 2017) points to the importance of assessing the biological relevance next to the statistical significance of the results, by integrating all available data: '… lack of statistical significance should not be the sole rationale for concluding a lack of exposure related effect, just as statistical significance should not be the sole justification for concluding on the occurrence of a treatment-related effect.'<br><br>If deviations from the control in the magnitude > 30–50 % are observed |

| | | |
|---|---|---|
| | | and are statistically significant, then the identification of an endpoint is considered unproblematic. If deviations from the control in the magnitude > 30–50 % are not statistically significant due to the poor statistical power of the assay but are considered biologically relevant following the evaluation as suggested above, then a weight-of-evidence approach might help to identify possible treatment-related effects (EFSA Scientific Committee, 2107). In this respect, evaluating the distribution patterns of the earthworms in the plots before application of the test substance can be helpful.<br><br>The following considerations can help the evaluation of the potential biological relevance of observed changes compared to control:<br><br>Is the distribution of abundance/biomass of aggregated data or on species level between the plots the same at the beginning and at test end?<br><br>Are observed differences possibly treatment-related?<br><br>Are decreases/increases to be observed through the study?<br><br>In the case of increases at test end: Are initial decreases in abundance or biomass followed by increases at test end, possibly indicating that the treatments are still not comparable to the controls (e.g. overcompensation)?<br><br>Does only one endpoint show effects or are more species or ecological groups affected? |

* It is recommended that layers of 1–2.5 and 2.5–5 are considered, which is in line with the EFSA Guidance on how to estimate predicted environmental concentration in soil (EFSA, 2017c).

## Proposal

- **Evaluation**

A structured evaluation of earthworm field studies as proposed by de Jong et al. (2006) is recommended and an extract of how this could be presented is shown below. The proposal reported in Table I2 is taken from the guidance by de Jong et al. (2006) with some modifications. The recommendations as indicated in Table I1 should be considered when assessing the reliability of each item and consequently the overall reliability of the study.

**Table I2**: Example of how to report and evaluate each relevant item of an earthworm field study

| Test item | Notes (questions to answer) | Reliability | Justification |
|---|---|---|---|
| 1. Substance | Was the representative formulation used?<br>If a vehicle is used, identity and concentration.<br>Substance used as reference item and at which dose. | 1-reliable<br>2-reliable with minor restrictions<br>3-reliable with major restrictions<br>4-not reliable | A justification of the reliability assessment should be provided |
| 4. Test design | Was the ISO guideline followed?<br>Plot size? | | |
| 6. Sampling | Was the earthworm sampling area as recommended?<br>Which sampling method was used? | | |

- **Earthworm sampling**

Among the different extraction methods for earthworms, the octet method has been shown to be inefficient, especially in dry condition and for anecic earthworms. Therefore, this method should preferably not be used especially as the only method used for the extraction.

- **Exposure in the test**

Although the ISO guideline only recommends verifying the application of the tested substance, it is suggested that the exposure is monitored over the duration of the test. Soil samples for residues analysis could be sampled at the same time points as the earthworm sampling. It is recommended that the concentration is measured in a way that it is made comparable to the PECsoil.

- **MDD analysis**

Although MDD is considered a valid concept for the post hoc evaluation of the statistical power of the test, before its routine use for the evaluation of earthworm field studies, additional guidance is needed on, for example, classes of MDD (%) and the minimum number of vulnerable taxa with an acceptable MDD.

# Appendix J – Expressing endpoints from Tier 1 tests and formulation tests (with one or more active substances) for unstable substances

**DISCLAIMER: This appendix represents a proposal from the Member States of the central zone. It was introduced at the meeting. It is included in this technical report because it provides recommendations that might be useful for other Member States with risk assessments for the authorisation of PPPs.**

The scope of this document is to reach a harmonised approach on:

- how to evaluate tests with one or more active substances that go beyond the recommendations from the EFSA peer review meeting on general recurring issues in ecotoxicology (EFSA, 2015);
- how to use this information in the context of the mixture toxicity risk assessment.

The Member States of the central zone are asked to reach an agreement concerning the expression of Tier 1 active substance and formulation endpoints. This position paper could be attached to the report of the EFSA peer review meeting on general recurring issues from October 2018.

Note: the EFSA Guidance on tiered risk assessment for edge-of-field surface waters (EFSA PPR Panel, 2013) is referred to as 'AGD'.

## 1 Background

At Tier 1, laboratory standard tests must be performed under standard (i.e. mostly worst case) exposure. Therefore, OECD guidelines recommend that the concentrations should be maintained and must be > 80 % and < 120 % of nominal at the end of the exposure period (or at the end of the renewal period for semi-static design).

If the concentration cannot be maintained (i.e. if the substance is dissipating 'fast'), the validity of the study should be questioned and the test may be rejected as highlighted during the EFSA peer review meeting on general recurring issues in ecotoxicology (EFSA, 2015).

During this EFSA peer review meeting, Member States agreed that in principle:

1) **Nominal concentrations** can be used to express the toxicity from any kind of test if the test concentrations were maintained at ± 20 % of the nominal at all times throughout the test including the study end sampling. Mean measured is also an option for this situation.

2) **Initial measured concentrations** can be used to express the toxicity from any kind of test if the initial test concentrations were below 80 % of the nominal and this concentration was maintained throughout the test (within ± 20 % of the initial) including the final sampling. Mean measured is also an option for this situation.

3) **Mean measured concentrations** must be used to express the toxicity from any kind of test when the test concentrations were not maintained within the range of ± 20 % of the nominal or initial measured, but significant concentrations of the test item were still present at the end of the exposure period (or at the end of the renewal period for semi-static design).

4) When the test concentrations were not maintained and significant residues were not present at the end of the exposure period (or at the end of the renewal period for semi-static design), the **validity of the study should be questioned**.

It was also pointed out that further clarifications should be provided in the AGD.

In practice (and not due to a causal relation), however, semi-static and/or flow-through design is rarely used for tests with:

- algae for which semi-static tests are very uncommon and flow-through tests not established in the regulatory context, due to the technical complexity when conducting the test

- formulated products with one or more active substance, especially for tests with algae.

This proposal addresses these issues. It especially considers the cases where the recovery of an active substance at the end of a test is < 80 % (i.e. the test substance is dissipating fast) and where requesting a new semi-static or flow-through test (as required by EFSA, 2015) may not be feasible or desirable (i.e. algae tests and vertebrate tests).

An adequate expression of the endpoint from formulated product tests is needed:
- for the purposes of classification and labelling, and
- as the basis for mixture toxicity assessment since it should enable an assessment of potential synergism or additive toxicity due to one or more co-formulants or additional active substances.

The proposed approach aims to serve both purposes.

Until a revision of the (EFSA PPR Panel, 2013) , this position paper is intended to fill the gap as an interim solution, i.e. for such cases where above-cited requirements 3 and 4 cannot be easily fulfilled and performing tests under semi-static or flow-through conditions are an issue.

## 2 Practical implications

### 2.1 Issues common to all Tier 1 tests

Tier 1 tests are performed under standard (i.e. mostly worst case) exposure conditions and ideally the exposure should be constant. Thus if the concentration of a test substance is known or expected to decrease by more than 20 % of the nominal or initially measured concentration during the test duration, an appropriate test design (i.e. semi-static or flow-through) must be chosen and adequate analytical measurements performed.

In practice, however, semi-static and/or flow-through design is rarely used for formulation (product) tests and generally not used for tests with algae (semi-static is very uncommon and flow-through not established in the regulatory context). In addition, in some cases of fast-dissipating substances, even semi-static test conditions may not guarantee sufficiently constant exposure.

This proposal addresses this issue and the cases where:

- the recovery of one or more substances at the end of a test is found or expected to be < 80 % (i.e. the test substance is dissipating fast and thus exposure conditions are regarded as sub-optimal), plus
- requesting a semi-static or flow-through test (as required by EFSA, 2015) may not be feasible (algae) or desirable (i.e. make best use of existing data for vertebrate and formulation tests), i.e.:

- Algae tests: these cannot be conducted in flow-through mode and only in rare cases in semi-static mode; please note that macrophyte tests are not of concern in this context since tests can be conducted under semi-static conditions. Please note that for algae tests performed with unstable substances, recovery of algae may have already occurred after 72 h, i.e. when $EC_{50}$ is usually calculated; consideration should be given to this aspect in the revision of the AGD, please see below[34].
- Vertebrate tests: existing tests and sufficient justification that fish are clearly not the most sensitive species (e.g. an endpoint several orders of magnitude higher than for other groups).
- Formulation tests: ideally, the tests should be carried out in flow-through mode (or under semi-static conditions) when one or more active ingredients are known to be dissipating; this is to achieve continuous exposure unless the active substances contained in the formulation are known to be sufficiently stable over the testing period (i.e. static design is justified). However, existing tests with formulations are often static tests. In principle, a new test should be requested if a reliable endpoint cannot be derived.

[34] For algae tests performed with unstable substances, consideration should be given to the duration selected for derivation of the endpoint in relation to the duration of the test (e.g. calculating $EC_{50}$ for 0 to 24 h, 0 to 48 h, 0 to 72 h, 0 to 96 h and select the lowest endpoint, while applying the validity criteria). This issue should better be tackled in the revision of the EFSA PPR Panel, 2013.

If a reliable endpoint for Tier 1 cannot be derived, the purpose/use of the endpoint in the regulatory context could be considered before requesting a new test (e.g. sensitivity of the tested species/group (i.e. if the group of organisms is clearly not most at risk), other information on its potential toxicity).

Please note the intention of the proposed approaches is *not* to decrease the conservativeness of Tier 1 test design guidelines.

### 2.2 Further considerations for formulation tests with more than one active substance

The initial assumption in the mixture toxicity assessment according to the (EFSA PPR Panel, 2013) is that only the active substances are responsible for the measured toxicity of the formulation.

In order to verify this assumption, product tests are required that allow for a meaningful comparison of calculated (i.e. expected) and measured (i.e. observed) toxicity. Consequently, tests conducted with formulated products should be highly comparable to the Tier 1 data for the active substances and therefore also be performed under standard (continuous) exposure conditions.

Ideally, in order to determine whether the concentration was maintained at > 80 % of nominal concentrations, analytical measurements are required for each active substance within the product, at least at the start and end of the test. If one or more active substances are not stable over the duration of the test, more frequent analytics are needed to describe the exposure. While this should be sought for any new product tests, there are product tests with more than one active substance that were performed without standard exposure and/or appropriate analytics and hence the following principles can be applied to such studies, too.

If analytical measurements indicate i) that concentrations have not been maintained or ii) are not available for all active substances, the procedure described below is proposed. It gives guidance on how to derive endpoints from such tests in order to make the best use of available data while being sufficiently protective within the risk assessment process.

Several cases can be distinguished. The current document presents the most common ones. Other cases can be added, if needed, in further revisions.

The main cases identified at the moment are:

**Case 1**: All active substances have been analytically measured. Please refer to Chapter 4.1.

**Case 2**: At least one active substance (but not necessarily all) has been analysed at least at the start and end of the test. However, the test may still be useful for the risk assessment under certain conditions. For further details, please refer to Chapter 4.2.

**Case 3**: Only initial measurements are given (for one or more active substance). This can impede meaningful comparison of calculated and measured toxicity. If such tests may still be useful is a case-by-case decision. See PseudoTox approach presented in Chapter 4.3.

## 3 Procedure for Tier 1 active substance tests and formulation tests with one active substance / active substance and monoformulation

In order to make best use of the available data, Tier 1 endpoints can be derived from calculated mean measured concentrations, if the analytical data provided are sufficient.

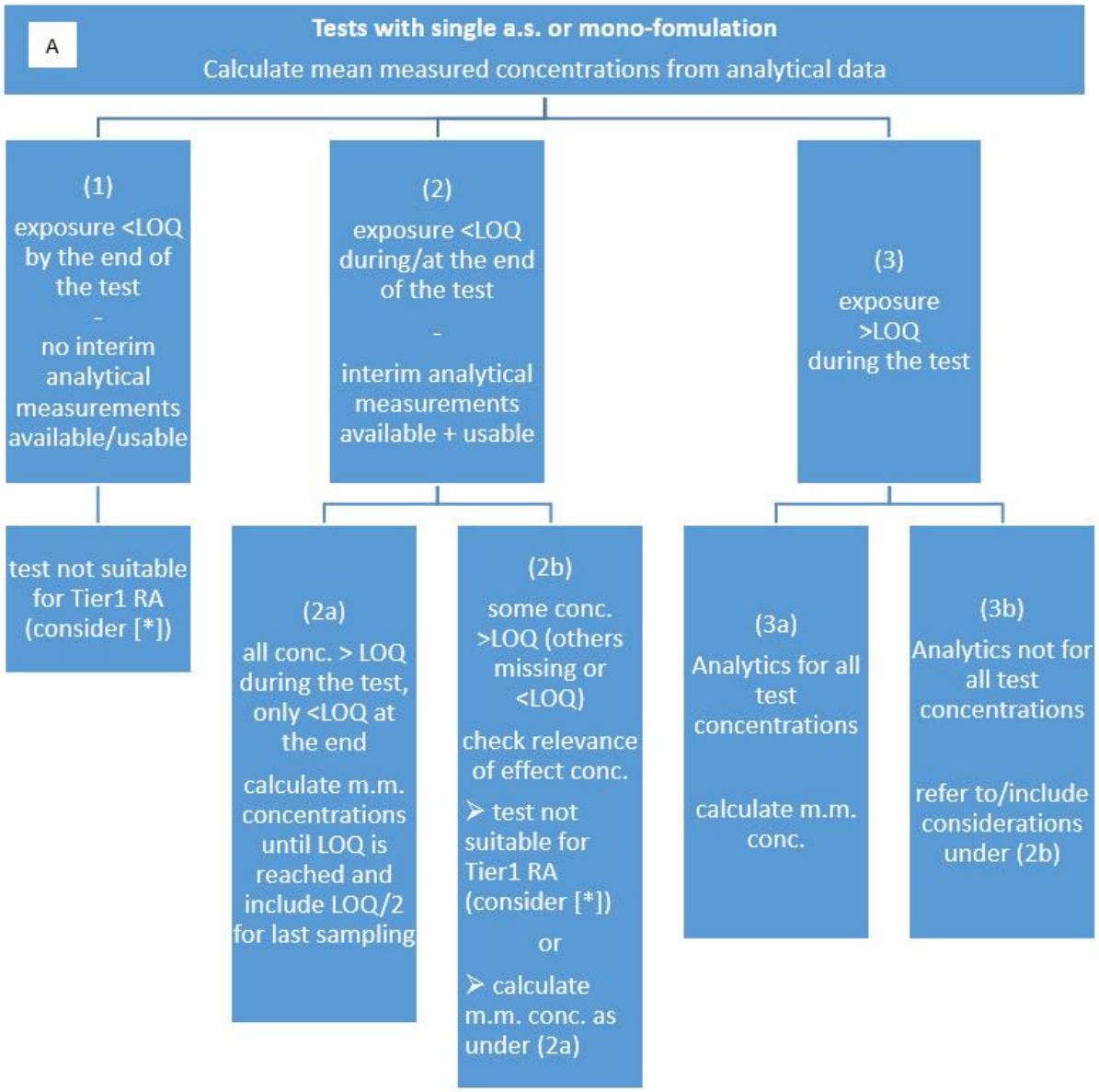
Depending on the stability and given analytical measurements, different cases arise. They are depicted in the decision schemes and explanations below.

| |
|---|
| Is the recovery of the tested substance showing constant exposure?<br>Yes<br>a) recovery within 80–120 % of nominal concentration: express endpoint as 'nominal' concentration<br>b) recovery within 80–120 % of initial measured concentration: express endpoint as 'initial measured' concentration.<br>For both a) and b), 'mean measured' concentration is also an option. |

No (recovery < 80 % or > 120 %): express endpoint as 'mean measured' concentration, See decision scheme A.

Decision scheme A, presented below, should be read in combination with the following section on 'Explanations'.

### 3.1 Decision scheme A (no constant exposure)




[*]: please refer to 3.2 (2).

## 3.2 Explanations of decision scheme A

The several cases in decision scheme A relate to the different types of analytical measurements available for tests with one active substance (and monoformulations). In general, tests with the active substances should be evaluated more strictly than tests with monoformulations since they should be considered as basic information, whereas formulation tests are often performed to identify effects of co-formulants. In addition, the sensitivity of the tested species/group (i.e. other information on its potential toxicity) should be considered, especially when deciding whether to request a new test.

(1) Cases when exposure was ≤ LOQ by the end of the test:
- analytical measurement shows exposure ≤ LOQ by the end of the test for (almost) all test concentrations and no interim analytical measurements are available;
- all interim analytical measurements are below the LOQ.

➢ **Test not suitable for Tier 1 RA → A new test with improved exposure design or improved analytical measurements is necessary.**

But consider [*]: Before requesting a new test, we recommend that the following cases are considered, for which a new test may not be justified:

(a) A new test with a monoformulation for a certain group of organisms might not be required if reliable endpoints of valid tests performed with the active substance indicate that another group of organisms is clearly more sensitive (i.e. a difference between the endpoints of a factor ≥ 10). This is particularly relevant for vertebrates (i.e. if invertebrates or algae are more sensitive) since vertebrate testing should be minimised for animal welfare reasons).

(b) For algae tests, if analytical interim measurements have been performed, a rejection is not adequate because the test design can hardly be improved (semi-static very uncommon and flow-through not established); in such cases calculate the test concentrations by, e.g., using SFO kinetics according to the estimated degradation/dissipation behaviour ($DT_{50}$ water). Then recalculate the endpoint. Also consider the possibility of adsorption to the increasing algal biomass. Please note that macrophytes are not an issue in this context (see Section 2.1).

(c) A new product test might not be required if it can be sufficiently justified that the effect is triggered by initial exposure levels within the test (supported by information from the active substance evaluation).

Please note that a test not suitable for Tier 1 risk assessment may, however, be useful in another part of the risk assessment (e.g. refined exposure test).

(2) Cases of sufficient analytical measurements with exposure ≤ LOQ reached during the test for (almost) all test concentrations and interim analytical and effect measurements are available:

(2a) All interim analytical measurements are above the LOQ for all test concentrations/only the last sampling (at the end of the test) resulted in concentration(s) < LOQ.

➢ make use of the available interim measurements until exposure ≤ LOQ and include LOQ/2 (proposal for pragmatic approach[35]) only for the last sampling time (see also point (3a)). Then recalculate the endpoint. The last analytics > LOQ should be from a time point not too far away from the last time point (e.g. for a 96 h fish test, last analytics at 48 h or 72h.

Example: In an acute fish test, intermediate analytics are available at 0 h, 24 h, 48 h, 72 h and 96 h. If analytics (around concentration delivering the endpoint) are > LOQ at 72 h and < LOQ at 96 h, it is recommended to build the geometric mean (geomean) between 0 and

[35] In the *Guidance Document On Aquatic Toxicity Testing Of Difficult Substances And Mixtures* (ENV/JM/MONO(2000)6) (OECD, 2018) it is stated that: 'In order to calculate a mean exposure concentration, the final concentration may be taken as the limit of detection for the method if the substance is not detected. When the substance is detected but not quantified, it may be a good practice to use half of the limit of quantification. Since there may be various methods for determining that, the method selected to determine mean measured concentrations should be made explicit in the reporting of test results.' As the limit of detection is not always reported within the relevant study protocol, the use of LOQ/2 is proposed for convenience, as a pragmatic approach.

96 h considering the measured analytical values up to 72 h and the LOQ/2 for 96 h (e.g. to avoid extra vertebrate testing).

(2b) Cases of interim analytical measurements are above the LOQ only for some test concentrations or not all test concentrations have been measured:

If there is a systematic lack of data, e.g. data are missing for all lower concentrations or all higher concentrations, and the non-detectable and/or non-measured concentration(s) is/are showing the relevant effects (e.g. close to $LC_{50}/EC_{50}$ for acute tests or close to $NOEC/EC_{10}$ for chronic tests):

➢ **Test not suitable for Tier 1 risk assessment → A new test is necessary with improved exposure design or improved analytical measurements.**

But consider [*]: Before requesting a new test, we recommend that the three cases are considered for which a new test may not be justified as described under point (1) above.

If there is no systematic lack of data, i.e. one (or some) lower and higher concentrations were analysed and these indicate that the dissipation patterns are similar (the standard deviation of the mean recovery rate is < 20 %, as it is the usual criteria for deviations of concentrations from nominal values), the following can be applied:

➢ Recalculate the missing concentrations assuming the same dissipation pattern as for the measured concentrations within the test. A mean recovery rate of the measured concentrations could be used to estimate the recovery for the missing concentrations. Thereby, the calculation would account for possible influences of the test system concerning the degradation/dissipation of the substance within the system. If this approach is applied this should always be pointed out and accompanied by the reason it is deemed suitable in the given case.

Calculate the geometric mean (geomean) for each test concentration based on the measured/calculated concentrations. Then recalculate the endpoint.

(3) Cases of analytic measurements with exposure > LOQ during the test:

(3a) analytic measurements for all test concentrations:

➢ To calculate the recovery of each test concentration expressed as geometric mean measured concentrations (no derivation of mean recovery), use these (geo)mean measured concentrations to calculate the endpoint (mean and confidence interval) expressed in mean measured concentration[36].

For semi-static test design: calculate for each test concentration the geomean within each renewal phase (e.g. geometric mean 0–2 days, geometric mean 2–4 days). Then, calculate the arithmetic mean of the geometric means of the renewal phases, e.g. case of one renewal: mean measured concentrations = (geometric mean 0–2 days + geometric mean 2–4 days)/2. Then recalculate the endpoint.

For formulation testing: Recalculate the test concentrations of the formulation based on the measured concentrations of the active substance.

Please also refer to the 'Guidance Document On Aquatic Toxicity Testing Of Difficult Substances And Mixtures' (ENV/JM/MONO(2000)6) (OECD 2018) and the corresponding OECD guideline for choosing the relevant calculation method (e.g. OECD GD 201 mentions 'geometric mean' for algae, OECD 211 mentions 'time-weighted mean' for daphnids).

[36] If the endpoint is not used or recommended for use in a risk assessment, a recalculation is not necessary.
If the endpoint is a 'higher than value' or only used to compare the active substance and monoformulation toxicity via the model deviation ratio, the following simplified calculation method may be used: the mean recovery rate could be calculated by averaging the recovery rates of all concentrations tested (expressed as (geo)mean measured concentrations). The mean recovery rate is then used to recalculate the endpoint (expressed in mean measured concentration).

(3b) analytic measurements not available for all test concentrations:
➢ Refer to (2b).

## 4 Procedure for formulation tests with more than one active substance

During the EFSA peer review meeting on general recurring issues in ecotoxicology (EFSA, 2015), definitions for the several cases were given (please refer to Chapter 1 – background).

The authors amended some of the definitions as follows for the purposes of this section on formulation with more than one active substance:

**Mean measured concentration**: the test concentrations of some or all active substance were not maintained within the range of ± 20 % of the nominal or initial measured (i.e. degradation or dissipation of some active substances needs to be considered) throughout the test duration, but significant concentrations of these active substances were still present at the end of the exposure period (or only the last measurement was found to be < LOQ). A geometric mean concentration for each active substance can be derived from these data.

**Peak measured concentration**: The measurements of active substance with dissipation from the water phase $DT_{50}$ ≤ 3 h often correspond to exposure ≤ LOQ at the time of measurement following the initial measurement (i.e. often approximately 24 h). This very fast dissipation of active substance needs to be considered and the endpoint expressed as peak concentration (corresponding to nominal or initially measured concentration only for the first measurement). Please note that this is not per se acceptable for a Tier 1 test design. Further aspects should be considered. For further information, please refer to Chapter 3.1.

The derivation of the product endpoint (sum of active substances) for a given test should always be justified and reported by taking the analytical measurements from the respective testing into account.

This section will not further consider the case of 'nominal' (i.e. all measured concentrations for all active substances within 80–120 %).

### 4.1 Case 1: All active substances have been analytically measured

Besides flow-through test design, this case (i.e. all active substances have been analytically measured) is the best situation in which to determine whether the endpoint of the product should be expressed as nominal, initial measured, peak measured or mean measured concentrations.

**Mean measured concentrations from a product with more than one active substance**: The measurements of some or all active substance are not within the range of 80–120 % of the nominal concentration (i.e. dissipation of some active substances needs to be considered). However, all substances could be analysed until the end of the test or only the last measurement was < LOQ.

In such cases, the following procedures might be applied:

**Option A** (preferred option because associated with fewer uncertainties) is carried out as follows:
(1) For each active substance, calculate the geometric mean concentration between the start and end of the test for each tested concentration; calculate the recovery rates at each tested concentration (geomean compared with nominal or initial measured).
(2) Sum up the new calculated geomean concentration levels for the active substances to derive the 'sum of active substances' per concentration level to calculate the endpoint in the following step.
(3) Calculate the endpoint (mean and confidence interval) based on the 'sum of active substances' geomean concentration levels using an appropriate statistical tool.

**Option B** (faster to calculate but associated with more uncertainties) is carried out as follows:

(1) For each active substance, calculate the geometric mean concentration between the start and end of the test for each tested concentration level; calculate the recovery rates at each tested concentration (geomean compared with nominal or initial measured).

(2) For each active substance, calculate the mean recovery rate and standard deviation, by considering the recovery rates for each concentration level as in (1). If the standard deviation is < 20 % of the mean recovery rate, the mean recovery rate can be used to recalculate the concentration levels for the respective active substance. If the difference is > 20 %, consider the plausibility and reliability of the analytical measurements (case-by-case decision).
(3) For each active substance, recalculate the mean measured concentration, based on the mean recovery rate.
(4) Sum up the new calculated concentration levels for the active substance to derive the mean 'sum of active substance' concentration levels.
(5) Recalculate the endpoint based on the recovery rates of the 'sum of active substances'.

Please refer to Example 1 (Example 1: all active substances measured / recovery not ± 20 % of nominal) in Chapter 5 for illustration.

For the purposes of comparing the measured (i.e. observed) and calculated (i.e. expected) toxicity of a formulated product, the resulting formulation toxicity data ($EC_{XPPP}$ in mg or µg formulation/L) have to be recalculated to the **sum of active substances** ($EC_{XPPP}$ in mg or µg sum of active substance/L) for use in mixture toxicity risk assessment. This calculation is based on the geomean measured concentrations of all active substances in the formulation test as total active substances.

**Peak measured concentration from a product with more than one active substance**: It might be justified to express the endpoint as peak measured concentration if the exposure ≤ LOQ after approximately 24 h for some or all active substances (i.e. $DT_{50}$ ≤ 3 h).

However, this is not an ideal Tier 1 test scenario and further information should be considered in the process of decision-making:

- It would apply to cases where requesting an ideal new test performed under flow-through conditions would not be wished (e.g. vertebrate tests for animal welfare reasons).
- Acceptable only if analytical sampling took place directly after the start of the experiments (exposure of individuals).
- If substances are not regarded as dissipating very fast, the test is not acceptable because it remains unclear when the concentrations dropped below the LOQ.
- For comparison at a later stage for a mixture toxicity assessment, only data from the same test design can be compared. Please also refer to Chapter 4.3 (Case 3: Only initial measurements are available).

## 4.2 Case 2: At least one active substance (but not all) has been measured at least at the start and the end of the test

For the calculation of the formulation endpoint for cases when at least one active substance has been analysed at the start and end of the test, generally the following is proposed:

- First, to assess the contribution of toxicity of each active substance in the formulation (using the principle of toxic units (TU)), in order to estimate whether there are significant contributions to toxicity of the formulation (see definition below).
- Second, to compare the dissipation patterns of the various active substances over a duration equivalent to the duration of the test (for active substance not measured, $DT_{50}$ values can be obtained from, e.g., the fate section or other relevant studies). This is in order to conclude whether one of the active substances is dissipating very fast[37] (see definition below).

Please refer also to decision scheme B and explanations below for a more detailed assessment.

Information on the toxicity and stability of active substances can be obtained from available ecotoxicological and fate studies (e.g. list of endpoints, DAR), read-across, etc.

[37] The criteria are adopted from the presentation 'Derivation of endpoints from studies with aquatic organisms' by Mathieu Pluijmen, Appendix 1 of Evaluation Manual for the Authorisation of Plant protection products according to Regulation (EC) No 1107/2009 (NL, 2017).

**Information needed**

1. **Based on the relative proportion of each active substance in the formulation and the Tier 1 toxicity endpoint of the individual active substance, is one of the active substances expected to not contribute significantly to the toxicity of the product?**

*Threshold of significant contribution to the overall toxicity:* see AGD, Chapter 10.3.7: less than 10 % of whole mixture toxicity/TU of the mixture.

Example: in a product containing the two active substances A and B, the contribution of B to the overall toxicity of the product is considered not significant when it contributes ≤ 10 % of the TU. In that case, A is considered to be driving the toxicity since it contributes ≥ 90 % of the TU (see Chapter 4.2.2 of this appendix).

2. **Based on available data, is one or more of the active substances expected to be dissipating very fast?**

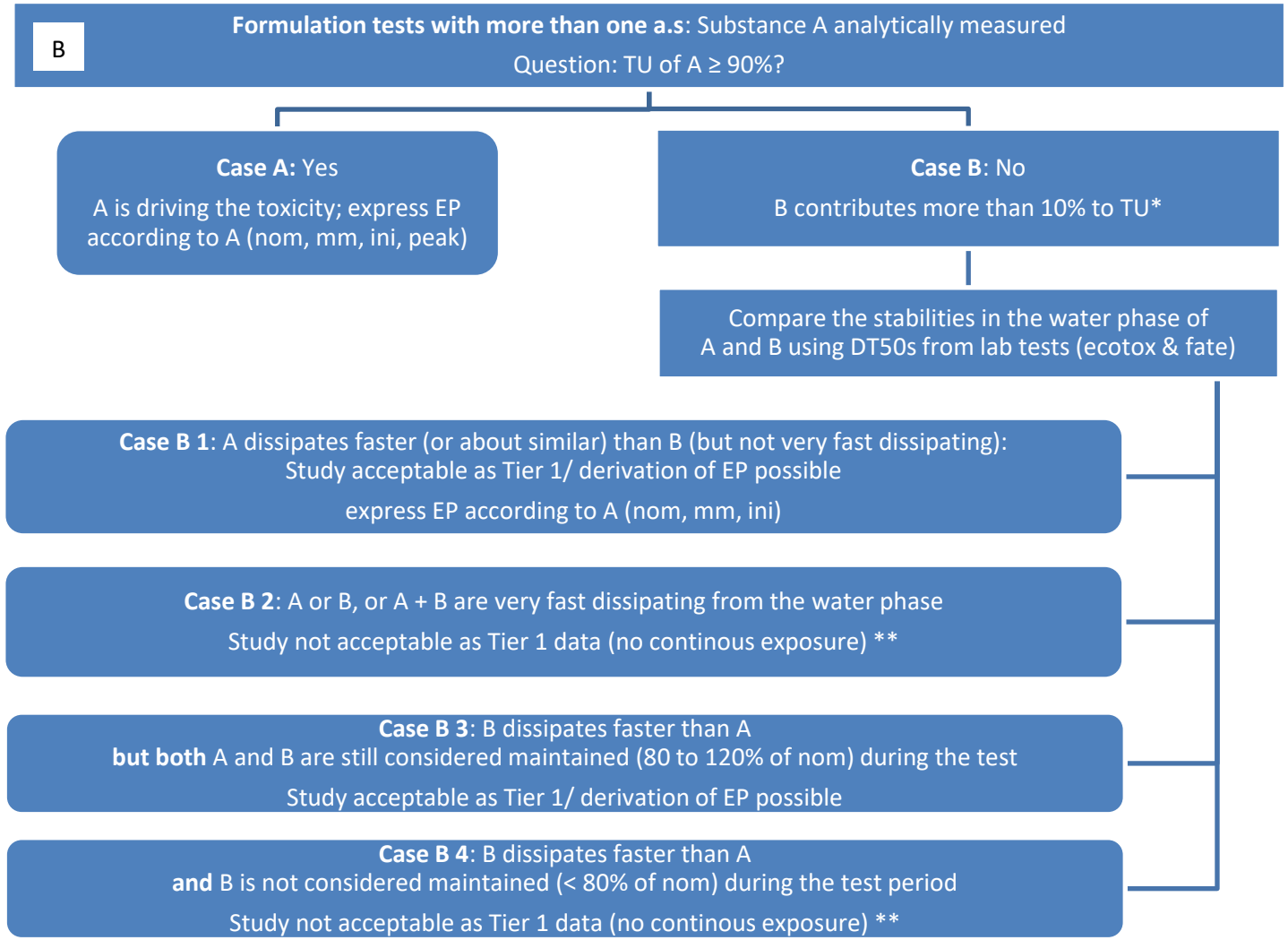
*Proposal for a threshold for the characteristic 'dissipating very fast'*:A substance is considered to be dissipating very fast when the expected dissipation time ($DT_{50}$) (i.e. based on information from other ecotox or fate tests) or measured $DT_{50}$ (i.e. in the given test) from the water phase is ≤ 3 h, which is equivalent to almost complete dissipation after 2 days (i.e. only 0.002 % left). This characteristic is indicated when the active substance in the given test is no longer measurable after 24 h at low nominal concentrations (<1 mg/L)[38]. In cases when sampling took place directly after the start of the experiment (exposure of organisms), using peak measured concentrations for the respective active substance may be acceptable; however, there are implications for the derivation of the product endpoint to consider.

Decision scheme B was developed on the basis of the information on toxicity and stability; it should be followed in order to decide whether a test concentration should be expressed in terms of nominal, initial measured, peak measured or mean measured concentration or if it is advisable to reject the test. Chapter 5 contains examples for illustration.

For the purpose of comparing measured (i.e. observed) and calculated (i.e. expected) toxicity of a formulated product, the resulting formulation toxicity data ($EC_{XPPP}$ in mg or µg formulation/L) have to be recalculated to the sum of active substances ($EC_{XPPP}$ in mg or µg sum of active substance/L) for use in a mixture toxicity risk assessment. This calculation is based on the known content of all active substances in the formulation. In the case of liquid PPP, the specific density might need further consideration. For the purpose of classification and labelling, note, however, that $EC_{XPPP}$ can be kept as mg or µg formulation/L.

[38] Threshold values taken from the same presentation of Mathieu Pluijmen (2015).

### 4.2.1 Decision scheme B



* When B is contributing > 10 % to TU, it is possible that in some cases it contributes > 90 %. In such cases, the derivation of a conservative endpoint might still be possible (i.e. if A dissipates faster than B). However, please consider carefully the possible implications on the mixture toxicity assessment. In such cases, requesting a new test will be a case-by-case decision. Justification for accepting or not accepting should always be included.

** If the non-measured substance or both substances are dissipating very fast (case B2), or if the non-measured substance is not considered maintained over the duration of the test (case B4), this test does not provide continuous exposure as generally requested for Tier 1. However, it may be possible to consider such a test in another context of the risk assessment (e.g. refined exposure test).

Such tests may be suitable to serve as substitute for Tier 1 tests with continuous exposure where requesting such a new test is not suitable (i.e. algae) or desirable (i.e. fish due to animal welfare reasons, especially in cases where vertebrates are not the most sensitive species). In such restricted cases, it needs to be carefully considered whether such data (without continuous exposure) could be adequate to perform a mixture risk assessment (e.g. screening for synergism).This would require a very thorough check and a comparison of data from single substance tests in a case-by-case assessment (e.g. when is the maximum effect reached, is there indication of recovery) and the reasoning should be presented (see also Chapter 4.3 Case 3).

If the assessment results in the conclusion that a new test is nevertheless required, a semi-static (in the case of primary producers) or flow-through set-up (for invertebrates and vertebrates) with analytics for both active substances should be recommended. However, a static test with appropriate analytics for the fastest dissipating substance may also be acceptable.

### 4.2.2 Explanations of decision scheme B

The focus of case 2 (i.e. at least one active substance, but not all, has been measured at the start and the end of the test) is to propose a systematic approach to decide whether a given product test is acceptable or not as the basis for the first step of the mixture toxicity risk assessment scheme.

In all explanations, active substance A is considered to be the active substance analytically measured.

**Step 1. Comparison of toxicity (pro rata)**

For tests with products containing several active substances, it is important to determine whether one active substance is driving the toxicity of the formulation. According to the aquatic guidance, one active substance is driving the toxicity if it contributes ≥ 90 % of the TU (see Chapter 10.3.7, EFSA PPR Panel, 2013).

In some cases, determining which active substance is driving the toxicity might be easy (e.g. if one active substance is more abundant and more toxic). In other cases, it is more difficult; for transparency reasons, it is thus suggested as a first step to always conduct a TU calculation to check and document whether there is a 'driver'.

As noted in the AGD, the TU is the ratio between the concentration (i.e. $c_i$) of a mixture component and its toxicological acute (e.g. $EC_{50}$) or chronic (e.g. long-term $EC_{10}$) endpoint. Toxic unit calculations should always be performed with toxicity data from the same species to avoid an influence of differing species sensitivity.

$TUi = \frac{ci}{ECxi}$ with $c_i$ being the amount of the active substance$_i$ [g/L] in the product and $EC_{Xi}$ being the Tier 1 endpoint of active substance$_i$.

In addition, the TU of a mixture has been defined as the sum of TU of each individual chemical of that mixture, i.e.

$$\sum_{i=1}^{n} TU = \sum_{i=1}^{n} \frac{ci}{ECxi}$$

Please note that this approach for identifying a driver of the toxicity is based on assuming concentration additivity of the individual active substance in the mixture which is usually the default assumption. Whether this assumption holds true or not refers to the mixture toxicity risk assessment (see Chapter 10.3 of the AGD).

The calculation of the TU will lead to one of the following results:

- A is contributing ≥ 90 % of TU (Please refer to Example 2 in Chapter 5)
- A is not contributing ≥ 90 % of TU (B contributes more than 10 % to the TU) (Please refer to Example 3 in Chapter 5).

**Step 2. Comparison of stability**

After calculation of the relative contribution of the active substance to the whole toxicity of the mixture (Step 1 'Comparison of toxicity (pro rata)'), it is suggested to compare the stability of the substances with respect to the dissipation/degradation behaviour in the relevant compartment, i.e. mostly the water phase of the test system; however, sediment could also be of importance in some cases (further considerations to be developed in the future).

To conduct this comparison, data from the environmental fate section should be consulted. By comparing the Deg/DissT50 of all active substances, it can be concluded which active substance is (a priori) dissipating/degrading most rapidly.

Special case of very fast dissipating substances: as presented above, active substances are considered to be 'dissipating very fast' when the expected dissipation time from the water phase is ≤ 3 h (i.e. analytics < LOQ after 24 h). In this case, endpoints can be expressed in terms of peak measured concentrations of active substance A if it is driving the toxicity. However, the protectiveness of an endpoint based on peak measured concentrations should always be carefully considered.

**General remark**: Please note that by recalculation of $EC_{XPPP}$ to the sum of active substances in terms of (geometric) mean measured concentrations of the faster degrading active substance, the MDR (model deviation ratio in the mixture toxicity assessment according to EFSA (2013)) will tend to increase and will thus indicate a potential synergism more frequently. In this case, careful consideration of the data quality is needed and possibly read-across for gathering information on the plausibility of the indicated synergism.

*Example for comparison of dissipation behaviour: A dissipating faster than substance B?*

Product containing two active substances (A and B); Summary of degradation in water/sediment system as reported in the 'fate and behaviour' part of the registration report:

| active substance | **DegT50** (whole system) | **DegT90** (whole system) |
|---|---|---|
| **A** | 69.9 days | 232.2 days |
| **B** | 5.3 days | 73 days |

➔ substance B dissipates faster than substance A.

## 4.3 Case 3: Only initial measurements are available

Generally, this option is not acceptable, and a new test should be required. Only in the case of product studies on fish from DAR (and RAR in the next renewal step) when fish is not the most

sensitive group could the following approach[39] be considered for use in mixture toxicity risk assessment; this approach is proposed in order to provide a suitably representative product toxicity endpoint for fish and prevent unnecessary further vertebrate testing.

Starting information needed:

Assuming that at least one active substance has been measured to confirm dosing in the product study and that studies of the same design (e.g. static) are available with the same active substance(s) as in the product, identify the following endpoints:

a. $EC_{XPPP}$ [mg sum a.s./L] from the product study, based on nominal or initially measured concentrations (but based only on measurements at test start).
b. Agreed Tier 1 endpoints for each active substance as used for the risk assessment of the respective active substance (based on the agreed appropriate expression as nominal, initial measured or mean measured concentrations according to the guidance and above illustrated principles for a single active substance).

   Concerning this point, three options are possible, i.e.:

   - 1 Revisit the underlying active substance toxicity studies and if not appropriately expressed, recalculate the endpoint to be correctly expressed before use in the pseudotox approach
   - 2 Do not revisit underlying active substance toxicity studies. Trust endpoint was appropriately expressed during the EU review
   - 3 Revisit underlying active substance toxicity studies and if not appropriately expressed, do not utilise the pseudotox approach to provide an estimated formulation toxicity endpoint.[40]

c. If, for any active substance, b. includes endpoints expressed in a different form from those of a., these endpoints for the respective tests should be revisited according to the principles illustrated above and additionally expressed in the same manner as a. (i.e. as initial measured or nominal concentrations – re-evaluation of agreed active substance studies might be needed).

Firstly, collect the Tier 1 endpoints for all active substances expressed as per the product endpoint, i.e. based on information from b. or, if b. is not delivering an endpoint expressed comparably to that of the product study (a.), based on information from c.

Next, the toxicity of the formulation assuming concentration additivity can be calculated (this is referred to as 'pseudotox' and the endpoint referred to as $EC_{x\text{-}Pseudo}$). This is done according to equation 13 (dealing with $EC_{X\ mix\text{-}CA}$) of the (EFSA PPR Panel, 2013) and will ensure a 'like-with-like' comparison of endpoints.

The resulting calculated toxicity can then be compared with the study-derived $EC_{XPPP}$ as described under a. This enables counter-checking calculated and measured $EC_X$ as a 'like-with-like' expression of endpoints is compared[41]. The method of doing so is described in equation 15 of the EFSA AGD ($EC_{X\text{-}pseudo}$ / $EC_{XPPP}$), with the result of this equation being the MDR. The result of this MDR calculation will dictate if/how estimated toxicity of the product can be used to assess the risk from the product and prevent the need for a further vertebrate study:

[39] Approach adopted from the UK's Health and Safety Executive draft paper *What to analyse in aquatic studies on plant protection products that contain more than one active substance – a discussion paper from the UK* (distributed for the Central Zone Harmonisation Workshop in Liverpool 2017).

[40] The Member States of the central zone expressed their position regarding these three options:
Option 1 was selected in majority: by 4 Member States (from which 1 country added some further considerations),
Option 2 was selected secondarily: by 2 Member States,
One Member State is not in favour of the pseudotox approach as a whole.

[41] Only studies conducted with the same design (e.g. semi-static) and with endpoints expressed in the same way (e.g. initial concentrations) can be used for this approach. The approach used for derivation of the endpoint for the active substance studies must be verified prior to applying this method, i.e. was it justified to use this method according to the principles described in this document?

1. If the calculated MDR is between 0.2 and 5, this indicates that the default assumption of concentration addition is reasonable[42]. In such instances there is then a need to recalculate the estimated $EC_{X\text{-}mix\text{-}CA}$ for the product, again using equation 13 of the AGD. However, the $EC_{X\text{-}mix\text{-}CA}$ now uses active substance endpoints expressed appropriately (i.e. information requirement b. detailed above).
   a) If the earlier-calculated MDR is in the range 0.2–1, then the predicted toxicity of the product – $EC_{X\text{-}mix\text{-}CA}$ – may be used directly in the risk assessment of the product. This is because the additive toxicity assumed in the calculation of this endpoint is established, and it would be expected that the predicted toxicity is conservative.
   b) If the earlier-calculated MDR is in the range > 1–5, then the predicted toxicity of the product – $EC_{X\text{-}mix\text{-}CA}$ – should be used in the risk assessment of the product, but a correction factor should be applied to this endpoint. This correction factor is based on the MDR value calculated (e.g. if $EC_{X\text{-}pseudo}$ / $EC_{XPPP}$ = 2.0, then $EC_{X\text{-}mix\text{-}CA}$ / 2 should be used in the product risk assessment). This is because, although it is established under a 'like-with-like' comparison of predicted and study-derived toxicity that an assumption of additive toxicity is appropriate, the indication is still that the product may be slightly more toxic than predicted. Therefore, in order to ensure a worst-case risk assessment (and protective associated risk mitigation recommendations) this predicted difference in toxicity should be accounted for.

2. If the calculated MDR is < 0.2, the product toxicity is established as less than additive. In such situations it would be suitably conservative, and also in line with the (EFSA PPR Panel, 2013) (Section 10.3.4) to use the predicted toxicity of the product – $EC_{X\text{-}mix\text{-}CA}$ calculated as described in 1, above – in the risk assessment of the product without a further correction factor.

3. If the calculated MDR is > 5, then more than additive (i.e. synergistic) toxicity is indicated. In this case it would not be appropriate to utilise the predicted product toxicity ($EC_{X\text{-}mix\text{-}CA}$) even with a correction factor, as the extent of increased toxicity over the assumed additive toxicity in this estimation is uncertain due to the deficient analytical results in the product toxicity study. In such cases the only feasible option would be to request a repeat study with appropriate analytical sampling and measurement.

## 5 Examples

### Example 1: all active substances measured / recovery not ± 20 % of nominal

If the option that is faster to calculate but associated with higher uncertainties was selected, the example is as follows:

A product A+B contains two active substances (a.s.) (100 g/L a.s. A + 10 g/L a.s. B). In a static test, both active substances have been analytically measured, with recoveries at test termination (2 days) ranging from about 60 to 80 % for active substance A and from about 30 to 50 % for active substance B.

Is the test acceptable for Tier 1 risk assessment? Since both active substances are measured and detectable during the test duration, the test is acceptable for Tier 1 risk assessment.

How should the endpoint of the product be expressed?

- geometric mean concentrations over the duration of the experiment (between test start (0 h) and end (2 d)) for each active substance and at each test concentration monitored for analytics were calculated (please see table below; in that example, all tested concentrations were monitored for analytics).

[42] The factor 5 corresponds to the MDR as introduced in Section 10.3.4 in the EFSA Guidance on tiered risk assessment for edge-of-field surface waters (EFSA PPR Panel, 2013).

- mean recovery rate and standard deviation were calculated for each active substance. In the example, the standard deviation is < 20 % of the mean recovery rate for each active substance.

**Table J1 - Example 1**: Overview of analytical measurements for active substances (a.s.) A and B, calculated geometric mean concentration over 2 days, recovery rate for each concentration level as well as mean recovery rate with standard deviation (absolute and in %)

| **Concentration level active substance A** | Nominal | | | | | Geomean Measured concentration | **Recovery rate** | | |
|---|---|---|---|---|---|---|---|---|---|
| | mg a.s./L | 0 h | 1 d | | 2 d | 0 – 2 d | **%** | | |
| 1 | 0.09 | 0.09 | 0.05 | | 0.03 | 0.051 | 57.00 | | |
| 2 | 0.19 | 0.19 | 0.12 | | 0.09 | 0.127 | 66.88 | | |
| 3 | 0.38 | 0.38 | 0.24 | | 0.17 | 0.249 | 65.62 | | |
| 4 | 0.75 | 0.75 | 0.5 | | 0.35 | 0.508 | 67.76 | | |
| 5 | 1.5 | 1.5 | 1.2 | | 0.9 | 1.174 | 78.30 | | |
| 6 | 3 | 3 | 2.5 | | 2.1 | 2.507 | 83.55 | **Standard deviation** | **In %** |
| | | | | | | **mean** | **69.85** | **9.54** | **13.7** |
| | | | | | | | | | |
| | | | | | | | | | |
| **Concentration level active substance B** | mg a.s./L | 0 h | 1 d | | 2 d | 0 - 2 d | **%** | | |
| 1 | 0.009 | 0.009 | 0.003 | | 0.001 | 0.003 | 33.33 | | |
| 2 | 0.019 | 0.019 | 0.01 | | 0.003 | 0.008 | 43.64 | | |
| 3 | 0.038 | 0.038 | 0.02 | | 0.008 | 0.018 | 48.03 | | |
| 4 | 0.075 | 0.075 | 0.041 | | 0.015 | 0.036 | 47.82 | | |
| 5 | 0.15 | 0.15 | 0.08 | | 0.007 | 0.044 | 29.20 | | |
| 6 | 0.3 | 0.3 | 0.1 | | 0.05 | 0.114 | 38.16 | **Standard deviation** | **In %** |
| | | | | | | **mean** | **40.03** | **7.79** | **19.5** |

- Mean measured concentrations for each active substance based on the mean recovery rate were recalculated (in this case, according to Option B, faster to calculate (see Section 4.1.)).
- Resulting concentrations for each tested concentration were added to derive mean 'sum of active substance' concentration levels (see Table J2 below).

**Table J2 - Example 1:** nominal and mean measured concentrations based on the mean recovery for active substances (a.s.) A and B, sum of active substance as contained in the product (nominal and based on mean measured concentrations)

| **Active substance A (mg a.s./L)** | | **Active substance B (mg a.s./L)** | | **Product (mg sum of a.s./L)** | | |
|---|---|---|---|---|---|---|
| Nominal | Recovered | Nominal | Recovered | Nominal | Recalculated based on recovery | Factor between nominal and recalculated concentrations |
| 0.09 | 0.063 | 0.009 | 0.0036 | 0.099 | **0.0666** | |
| 0.19 | 0.133 | 0.019 | 0.0076 | 0.209 | **0.1406** | |
| 0.38 | 0.266 | 0.038 | 0.0152 | 0.418 | **0.2812** | |
| 0.75 | 0.525 | 0.075 | 0.03 | 0.825 | **0.555** | |
| 1.5 | 1.05 | 0.15 | 0.06 | 1.65 | **1.11** | |
| 3 | 2.1 | 0.3 | 0.12 | 3.3 | **2.22** | **0.6727** |

- With the recalculated concentrations of the product (mg sum of a.s./L), the relevant endpoint for this study can be recalculated.
  e.g. endpoint nominal = 0.65 mg sum of a.s./L; endpoint based on mean recovery = 0.44 mg sum of a.s./L (recalculated with respect to the factor of 0.67 between nominal and recalculated concentrations).

**Example 2: only one active substance measured / TU of active substance A ≥ 90 % (case A of scheme 4.2.1)**

A product A+B contains two active substances. In a static test, only A has been analytically measured, with recoveries ranging from 40 to 60 % at test termination.

Is the test acceptable for Tier 1 risk assessment?

How should the endpoint of the product be expressed?

- The given formulation A+B contains 200 g A/L and 150 g B/L.
- The respective agreed Tier 1 endpoints for the individual active substance are:
  A: $EC_{50}$ = 0.02 mg A/L (nominal, derived from a flow-through test)
  B: $EC_{50}$ = 1.0 mg B/L (mm, derived from a static test).

Step 1. Comparison of toxicity (pro rata)

| **Active substance (a.s.)** | **Endpoint [mg/L]** | **Amount of a.s. in product [g/L]** | **Proportion in product [%]** | **Toxic units** | **Relative toxic units [% TU]** |
|---|---|---|---|---|---|
| **A** | 0.02 | 200 | 0.571 | 10 000 | 98.52 |
| **B** | 1 | 150 | 0.429 | 150 | 1.48 |

➔ driver identified (relative contribution of active substance A to the overall toxicity is ≥ 90 %).

Conclusion:

- substance A has been analytically measured
- A is contributing ≥ 90 % of TU.

➔ Study acceptable for Tier 1 risk assessment. Since active substance A is driving the toxicity of the product, the endpoint of the product should be expressed according to the measurements of active

substance A (case A in decision scheme 4.2.1) (in this example expressed in mean measured concentration since the recoveries of analytics for A range from 40 to 60 % at test termination).

**Example 3: only one active substance measured / TU of active substance A < 90 % (case B of scheme 4.2.1)**

A product A+B contains two active substances. In a static test, only A has been analytically measured.

Is the test acceptable for Tier 1 risk assessment?

How should the endpoint of the product be expressed?

- The given product A+B contains 200 g A/L and 150 g B/L.
- The respective agreed Tier 1 endpoints for the individual active substances are:
  A: $EC_{50}$ = 0.4 mg A/L (nominal, derived from a static test)
  B: $EC_{50}$ = 0.2 mg B/L (nominal, derived from a flow-through test).

Step 1. Comparison of toxicity (pro rata)

| Active substance | Endpoint [mg/L] | Amount of active substance in product [g/L] | Proportion in product [%] | Toxic units | Relative toxic units [% TU] |
|---|---|---|---|---|---|
| **A** | 0.4 | 200 | 0.571 | 500 | 40 |
| **B** | 0.2 | 150 | 0.429 | 750 | 60 |

➔ no driver identified (no relative contribution to the overall toxicity is ≥ 90 %).

Example 3, Case B1: Step 2. Comparison of stability
The available Tier 1 data (including additional studies and environmental fate data) indicate that A dissipates faster than B (DissT50 water of active substance A << than for B).
Conclusion:

- substance A has been analytically measured
- A is not contributing ≥ 90 % of TU (B contributes more than 10 % of TU)
- A dissipates faster than B.

➔ The study is acceptable for Tier 1 risk assessment since A dissipates faster than B. The product endpoint should be derived according to the analytical measurements of active substance A. This leads to a conservative endpoint, implications for mixture toxicity risk assessment need to be considered.

Example 3, Case B2: Step 2. Comparison of stability
The available Tier 1 data (including additional studies and environmental fate data) indicate that B has a DissT50 of < 24 h (B supposed to be very fast dissipating). A dissipates slower than B.
Conclusion:

- substance A has been analytically measured
- A is not contributing ≥ 90 % of TU (B contributes more than 10 % of TU)
- A dissipates slower than B (B dissipating very fast).

➔ The study is not suitable for Tier 1 risk assessment since B contributes more than 10 % to the whole toxicity of the mixture and is clearly less stable than the analytically measured substance A. However, the study might be acceptable for higher tier risk assessment (e.g. Tier 2C).

Example 3, Case B3: Step 2. Comparison of stability
The available Tier 1 data (including additional studies and environmental fate data) indicate that B dissipates faster than A but A and B can both be considered stable (recovery > 80 %) over the duration of the specific test (e.g. *Daphnia* acute test (2 d) or *Lemna* test in semi-static test design).
Conclusion:

- substance A has been analytically measured
- A is not contributing ≥ 90 % of TU (B contributes more than 10 % of TU)
- A dissipates slower than B but both are considered stable according to the test duration.

➔ The study is acceptable for Tier 1 risk assessment since A and B can both be considered as stable according to the specific test duration. The product endpoint should be derived according to the analytical measurements of active substance A (nominal or initial measured).

**Example 4: Only initial measurements are available**

A product A+B contains two active substances. In an acute toxicity test with fish under static conditions only A has been analytically measured, and then only at test initiation, with recoveries ranging from 70 to 99 %. Is the test acceptable for Tier 1 risk assessment? How should an endpoint for the product be expressed and used?

- The given formulation A+B contains 200 g A/L and 150 g B/L, product density = 1.0
- The acute fish study report expressed the product toxicity as LC50 = 7.5 mg product/L, based on initial measured concentrations.
- The respective Tier 1 endpoints for the individual active substance are:
  A: LC50 = 0.5 mg A/L (nominal, derived from a static test)
  B: LC50 = 2.0 mg B/L (nominal, derived from a static test).

Step 1. Collect all endpoints and express in a 'like-with-like' manner.

Active A = 0.5 mg a.s./L (nominal, static test) → 0.55 mg a.s./L (initial measured, static test)*
Active B = 2.0 mg a.s./L (nominal, static test) → 1.68 mg a.s./L (initial measured, static test)*
**recalculated with reference to study summary in DARs*
Please note that endpoints expressed as 'initial measured' concentration refer to initial measured concentration but based only on measurements at test start, as indicated in Section 4.3.

LC50 for product = 7.5 mg product/L (initial measured, static test) + considering 350 g sum a.s./L product→ 2.625 mg sum a.s./L (initial measured, static test).

Step 2. Estimate the toxicity of the formulated product assuming concentration additivity according to equation 13 of the aquatic guidance document (EFSA PPR Panel, 2013).

| **Active** | **LC50** | **proportion w/w** | **p1/LC50** |
|---|---|---|---|
| A | 0.55 | 0.571 | 1.039 |
| B | 1.68 | 0.429 | 0.255 |
| Sum | | | 1.294 |
| **LC50 – pseudo (mg a.s./L)** | | | **0.773** |

Step 3. Calculate the MDR by dividing the estimated toxicity (LC50-pseudo) by the study-derived product toxicity to see if additive, less-than-additive or more-than-additive toxicity is indicated. Ensure the same expression of endpoint (sum active substance) is used.

| | | |
|---|---|---|
| Formulation toxicity (pseudo): calculated | 0.773 | (Expressed in total active substance) |
| Formulation toxicity: measured | 2.625 | (Expressed in total active substance) |
| **MDR** | **0.294** | |

| MDR key | | |
|---|---|---|
| | < 0.2 | Less than additive |
| | 0.2–5 | Additive |
| | > 5 | More than additive (synergistic) |

Step 4. Decide if the estimated product endpoint based on assumption of additive toxicity can be used in risk assessment.

Calculated MDR supports additive toxicity → can be used in risk assessment. Recalculate estimated product toxicity using equation 13 of the (EFSA PPR Panel, 2013) but using individual active substance endpoints expressed appropriately.

Active A: LC50 = 0.5 mg a.s./L (nominal, static test)
Active B: LC50 = 2.0 mg a.s./L (nominal, static test)

| Active | LC50 | proportion w/w | p1/LC50 |
|---|---|---|---|
| Active 1 | 0.5 | 0.571 | 1.143 |
| Active 2 | 2.0 | 0.429 | 0.214 |
| Sum | | | 1.357 |
| **LC50 mix-ca (mg a.s./L)** | | | **0.737** |

Step 5. Decide if an additional correction factor is required to be applied to the estimated product endpoint.

As the earlier-calculated MDR was in the range 0.2–1 no correction factor is required to be applied to the LC50 mix-ca, as the assumption of additive toxicity is shown to be on the conservative side (i.e. MDR < 1).

**Product endpoint for use in risk assessment = 0.737 mg a.s./L**

Conclusion:

- Only substance A has been analytically measured.
- The full exposure duration has not been analytically verified (only initial measurements taken).
- Comparison of the study endpoint to a 'like-with-like' endpoint supports additive toxicity as being both appropriate and suitably conservative.
- An estimated product LC50 based on accurately expressed individual active substance endpoints may be used in the risk assessment for the product, without adjustment of the endpoint using a correction factor.
- Such an approach is only relevant for vertebrate studies, where all utilised studies are of the same design (e.g. static, semi-static, flow-through).

**Example 5: Only initial measurements are available**

A product A+B contains two active substances. In an acute toxicity test with fish under static conditions only A has been analytically measured, and then only at test initiation, with recoveries ranging from 70 to 99 %. Is the test acceptable for Tier 1 risk assessment? How should an endpoint for the product be expressed and used?

- The given formulation A+B contains 200 g A/L and 150 g B/L, product density = 1.0
- The acute fish study report expressed the product toxicity as LC50 = 1.2 mg product/L, based on initial measured concentrations.
- The respective Tier 1 endpoints for the individual active substance are:
  A: LC50 = 0.5 mg A/L (nominal, derived from a static test)
  B: LC50 = 2.0 mg B/L (nominal, derived from a static test)

Step 1. Collect all endpoints and express in a 'like-with-like' manner.

Active A = 0.5 mg a.s./L (nominal, static test) → 0.55 mg a.s./L (initial measured, static test)*
Active B = 2.0 mg a.s./L (nominal, static test) → 1.68 mg a.s./L (initial measured, static test)*

**recalculated with reference to study summary in DARs*
Please note that endpoints expressed as 'initial measured' concentration refer to initial measured concentration, but based only on measurements at test start, as indicated in Chapter 4.3.

LC50 for Product = 1.2 mg product/L (initial measured, static test) + considering 350 g sum a.s./L product
→ 0.42 mg sum a.s./L (initial measured, static test)

Step 2. Estimate the toxicity of the formulated product assuming concentration additivity according to equation 13 of the (EFSA PPR Panel, 2013).

| **Active** | **LC50** | **Proportion w/w** | **p1/$EC_{50}$** |
|---|---|---|---|
| A | 0.55 | 0.571 | 1.038181818 |
| B | 1.68 | 0.429 | 0.255357143 |
| Sum | | | 1.293538961 |
| **LC50 – pseudo (mg a.s./L)** | | | **0.773** |

Step 3. Calculate the MDR by dividing the estimated toxicity (LC50-pseudo) by the study-derived product toxicity to see if additive, less-than-additive or more-than-additive toxicity is indicated. Ensure the same expression of the endpoint (sum active substance) is used.

| | | |
|---|---|---|
| Formulation toxicity: calculated | 0.773 | (Expressed in total active substance) |
| Formulation toxicity: measured | 0.420 | (Expressed in total active substance) |
| **MDR** | **1.840** | |

Key < 0.2 Less than additive
0.2–5 Additive
> 5 More than additive (synergistic)

Step 4. Decide if the estimated product endpoint based on assumption of additive toxicity can be used in the risk assessment.

Calculated MDR supports additive toxicity → can be used in the risk assessment. Recalculate estimated product toxicity using equation 13 of the (EFSA PPR Panel, 2013) but using individual active substance endpoints expressed appropriately.

Active A: LC50 = 0.5 mg a.s./L (nominal, static test)
Active B: LC50 = 2.0 mg a.s./L (nominal, static test)

| **Active** | **$EC_{50}$** | **Proportion w/w** | **p1/$EC_{50}$** |
|---|---|---|---|
| Active 1 | 0.5 | 0.571428571 | 1.142857143 |
| Active 2 | 2.0 | 0.428571429 | 0.214285714 |
| Sum | | | 1.357142857 |
| **LC50 mix-ca (mg a.s./L)** | | | **0.737** |

Step 5. Decide whether an additional correction factor is required to be applied to the estimated product endpoint.

As the earlier-calculated MDR was in the range 1–5 a correction factor is required to be applied to the LC50 mix-ca, as the assumption of additive toxicity is shown to be slightly under-representative of measured product toxicity (i.e. MDR > 1 but in the range 1–5). As such the LC50 mix-ca should be divided by the MDR of 1.84 to account for this.
**Product endpoint for use in risk assessment = 0.737 / 1.84 = 0.401 mg a.s./L**

Conclusion:

- Only substance A has been analytically measured.
- The full exposure duration has not been analytically verified (only initial measurements taken).
- Comparison of the study endpoint with a 'like-with-like' endpoint supports additive toxicity as being supported, but slightly under-representative of indicated product toxicity.
- An estimated product LC50 based on accurately expressed individual active substance endpoints may be used in the risk assessment for the product, but with adjustment of the endpoint using a correction factor equivalent to the calculated MDR.
- Such an approach is only relevant for vertebrate studies, where all utilised studies are of the same design (e.g. static, semi-static, flow-through).