PSN, 12 May 2026

# CATS IN ECOTOXICOLOGY

Laura Villamar

Ecotoxicology team – PLANTS Unit

EUROPEAN FOOD SAFETY AUTHORITY

# CATS IN ECOTOXICOLOGY - INITIAL PROJECT


- The project OC/EFSA/PREV/2020/01 was launched and evaluated in 2020 and started in July 2021 – ran for 1 year
- Project developed Critical Appraisal Tools (CATs) for the evaluation of certain types of studies commonly used in ecotoxicological evaluation of active substances
- Developed by a consortium of well-known risk assessment agencies and other supporting bodies
- Proposal for CATs were published in 2023
- Implementation in the regulatory environment was needed

National Institute for Public Health and the Environment
Ministry of Health, Welfare and Sport

anses

WAGENINGEN UNIVERSITY & RESEARCH

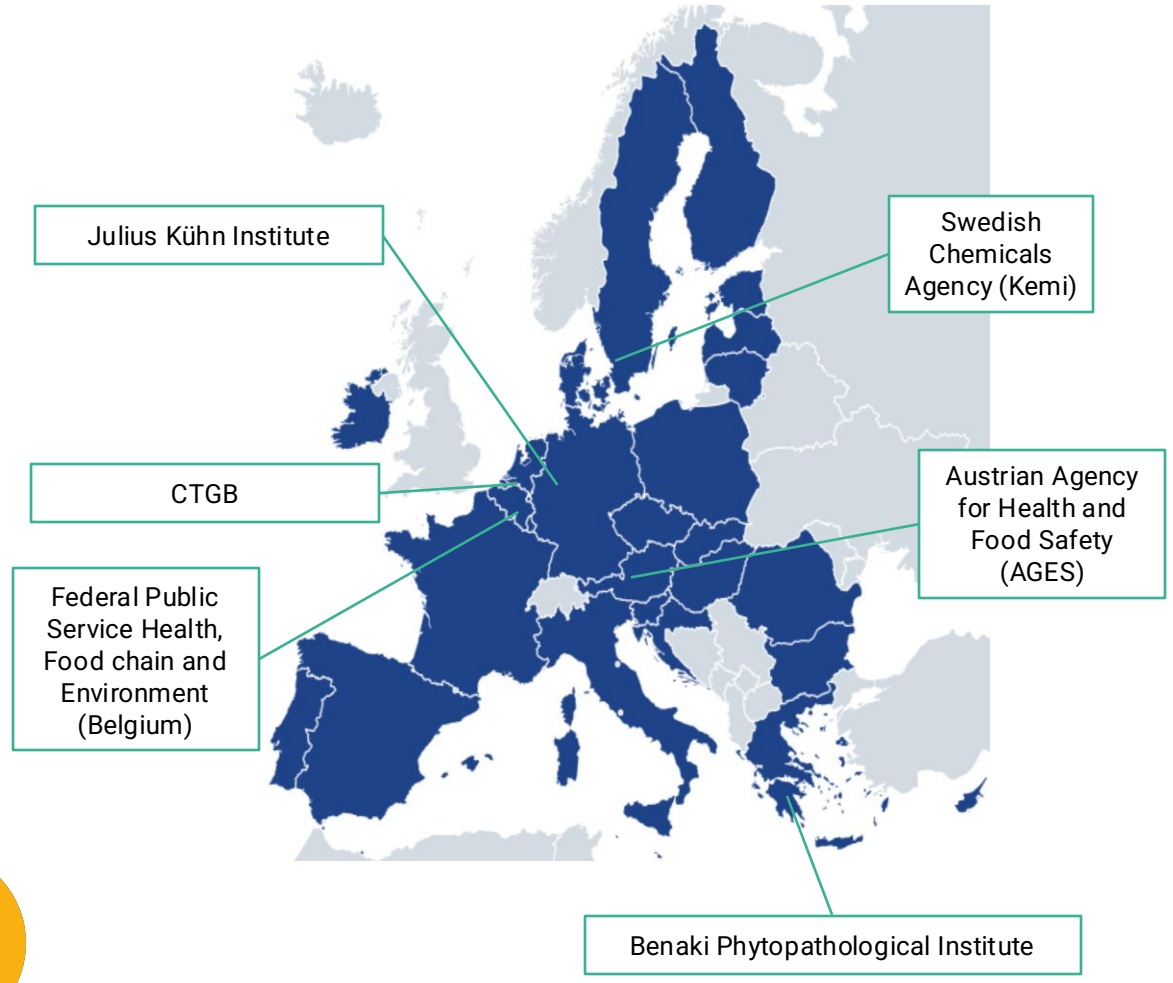
Umwelt Bundesamt

National institute of Public Health and Environment (RIVM) ( coordinator)
Wageningen Environmental Research ( WENR)
Agency for food, Environmental and Occupational Health and Safety ( ANSES)
German Environmental Agency/ Umweltbundesamt (UBA)



teamplantsecotoxicology@efsa.europa.eu

# WHY WAS TESTING NEEDED?




- Following the publication of the report, testing was considered necessary for transparency and to ensure the adequacy of the documents for the regulatory EU process
- Engage also with ecotox experts from MSs
- A call for expression of interest was launched, and the testing group was formed
- Testing was conducted for 6 months and experts tested real-case studies and provide feedback
- Every feedback was discussed during the internal meetings, and controversial or difficult points were proposed for a discussing in the General meeting
- As result of the discussion and agreement in the general meeting, EFSA issues a TR and the updated CATs (*in press*)
- Implementation was agreed by MSs
- Next engagement with PAFF and PSN

# WHY CRITICAL APPRAISAL TOOLS (CATS) NEEDED?

Standard and not-standardized studies currently co-exist in the pesticides' dossiers in ecotoxicology. Particularly non-standardized studies, evaluation can be more complex and of difficult interpretation **depending in the level of expertise of the evaluator**

Critical Appraisal Tools (CATs) provide a **structured approach to**

**- assess the reliability and relevance of individual studies,**

**- make the assessment more transparent,**

**- and increase consistency among different active substances.**

For each study type, an excel-based CAT was developed and it is accompanied by a **handbook** for guiding the assessor in the evaluation of each criterion

# HOW AN EFSA CAT LOOKS LIKE?

**Critical Appraisal Tool (CAT) - Reliability and relevance assessment of mesocosm studies with aquatic organisms**

*Adapted from Lahr et al., 2023. See also EFSA's general meeting PREV 32 (EFSA, 2026).*

Example of Excel CAT

| | |
|---|---|
| **Evaluation details** | |
| **Study information** | Add full study reference |
| **Evaluator information** | Add evaluator or institution details |
| **Date of evaluation** | Add date of evaluation |
| **Study conducted under GLP?** | [Select one] |

The cells colored in blue and green below indicate the reliability and relevance assessments generated after the assessment tool has been completed. They will be automatically filled once the evaluation is concluded and the evaluator refresh the data to get the final classification. If the risk assessor disagrees with the score provided by the tool, a robust justificat must be entered in cells F15 and F18.

| Reliability assessment | Reliability score generated by the tool | Reliability based on experts' judgement | Justification of selection of different reliability score |
|---|---|---|---|
| | Moderate reliability | [Select one] | *Justification to be given by the risk assessor* |

| Relevance assessment | Relevance score generated by the tool | Relevance based on expert's judgement | Justification of selection of different relevance score |
|---|---|---|---|
| [Description of the assessment goal (needed for relevance evaluation) (to be included in I64)] | Low relevance | [Select one] | *Justification to be given by the risk assessor* |

**General considerations/instructions**

Before evaluating the test, please check the physicochemical characteristics of your compound (handbooks/general sources). What is the solubility, log KOW, pKa? Is the compound volatile? Does it hydrolyse, photolyse, etc.? What is the mode of action (MoA) of the active? What are relevant metabolites and their properties? If any properties could influence study setup or results these should be mentioned in the rationale box in row 27 (cell I27) for the general information.

Before evaluating the study, please check whether the test is performed to a guideline method or a guidance document. **This tool needs to be used with the corresponding handbook**. Mark your evaluation per criterion with an 'x' in the respective box. An explanation of the reasons for the classification should be added to the rationale box.

If the same study is used in different regulatory frameworks, the reliability assessment can stand the same; however, the relevance assessment needs to be adapted to the pertinent risk assessment situation.

**Reliability assessment of mesocosm studies with aquatic organisms**

| Criterion code | | YES, criterion fulfilled | Yes, but criterion partly fulfilled | NO, criterion not fulfilled | Criterion not reported | Criterion not applicable | Type of criteria | Rationale |
|---|---|---|---|---|---|---|---|---|
| | **General information** | | | | | | | |
| G-1 | Is a study plan followed? Are deviations from the study plan reported and appropriately justified? | | x | | | | Other criteria | |
| | **Validity criteria and control response** | | | | | | | |
| V-1 | Are appropriate negative controls included and does the community present in the controls remain representative for a realistic freshwater community (including sensitive and/or vulnerable species) for edge-of-field surface waters during the whole test duration? | x | | | | | Key | |
| V-2 | Are clear effects seen at the highest treatment level? | x | | | | | Key | |
| | **Test item** | | | | | | | |
| T-1 | Is the test substance or test item clearly identified? Is the formulation type and amount of active substance reported? | x | | | | | Key | |
| T-2 | Are test results reported for the appropriate form of the compound? | x | | | | | Key | |
| T-3 | Are test results reported for the relevant compounds (i.e. active substance(s) and/or metabolites if pertinent)? | x | | | | | Key | |
| | **Test location** | | | | | | | |
| L-1 | Are the geographical location and source of environmental materials (from where water, sediments and organisms are originated, i.e., site of origin) specified, and uncontaminated? | | x | | | | Key | |
| L-2 | Has the study author justified that measures were taken to ensure minimal stress to the test organisms during the pre-experimental period? | x | | | | | Other criteria | |
| | **Biological system** | | | | | | | |

| | |
|---|---|
| T-1 | Is the test substance or test item clearly identified? Is the formulation type and amount of active substance reported? |
| T-2 | Are test results reported for the appropriate form of the compound? |
| T-3 | Are test results reported for the relevant compounds (i.e. active substance(s) and/or metabolites if pertinent)? |

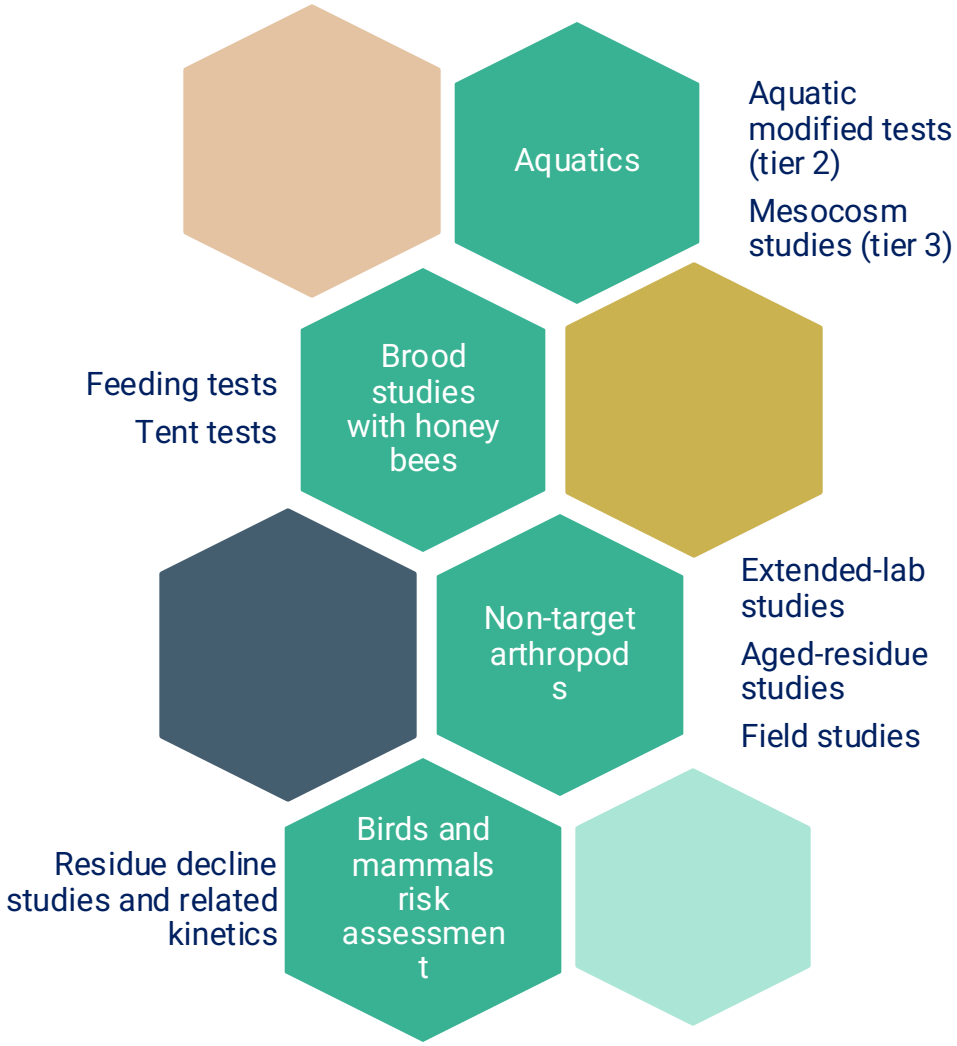
Example of the handbook

**Criterion T-3: Are test results reported for the relevant compounds (i.e., active substance (s) and metabolite(s))?**

| | |
|---|---|
| **Explanation** | If the test substance is subject to biotic or abiotic degradation during the test, metabolites may be formed, and their accurate analytical measurements might be required.<br>This is especially the case when metabolites are contributing to the total toxicity in the test. of equal toxicity (endpoint value within a factor 3 of parent endpoint) or higher toxicity compared to parent A<br>Information about the toxicity of the parent and its metabolite(s) can be deduced from lower tiers tests. The common practice is to consider a metabolite of i) equal toxicity to the parent when the endpoint value is within a factor 3 of the parent endpoint, and ii) higher toxicity when the endpoint value deviates more than a factor 3 from the parent endpoint. Also a metabolite can be considered relevant if the lower tier risk assessment indicated an unacceptable risk. It is thus important to assess whether the analytical measurements of the metabolite(s) - as for the parent compound - are sufficient for a reliable evaluation of the fate of the substance in the mesocosm (see E-6) since such metabolites are potentially relevant. |
| **Appraisal** | **Rationale for appraisal** |
| **Criterion fulfilled** | Potentially relevant metabolites, i.e., metabolites substantially contributing to the overall toxicity, are measured and quantified. |
| **Criterion partially fulfilled** | Not applicable. |
| **Criterion not fulfilled** | Although toxicity data give indication of metabolites contributing to the overall toxicity, this is not appropriately considered in the mesocosm study. |
| **Justification** | The PPR Panel (EFSA, 2013) has developed an assessment scheme for risk assessment of metabolites for lower tier assessment. For the assessment of the metabolite, the applicant has to provide a reasoned case whether the molecule contains a toxophore or if it has been lost following transformation. See section 10.2.2 for further guidance (EFSA, 2013). In lower tier risk assessment, if there are indications of toxic metabolites, this may require considering a mixture toxicity approach (See section 10.3 for further guidance (EFSA, 2013)). But for mesocosms, the endpoint(s) derived for the populations at risk reflects the integrated toxicity to which the individuals of this population were exposed. If lower tier risk has indicated a risk due to metabolites, then considering the exposure to the metabolite (maximum predicted exposure concentrations and predicted exposure profiles) become important for the subsequent risk assessment (i.e., considering the predicted exposure profiles of the metabolites in addition to those of the parent compound, see envelope curve concept in R9). |

# TYPE OF CATS AND APPLICABILITY DOMAIN

Study types and ecotoxicology applicability domain covered by the EFSA CATs



# THE CRED METHODOLOGY

- CRED methodology more commonly used in Ecotoxicology (Moermond, *et al.* 2016)
- Two main groups of criteria are distinguished, for reliability of the study and for its relevance.
- **Reliability** is defined as the inherent quality of an effect value in a test report or publication relating to (1) a clearly described experimental design to allow for the study to be repeated independently, (2) the way the experimental procedures were performed and (3) the reporting of the results to provide evidence of the reproducibility and accuracy of the findings (Lahr et al., 2023). It is important to note that the reliability is not dependent on the goal or regulatory framework of a study.
- **Relevance** is defined as the extent to which data and tests are appropriate to answer a specified question from a regulatory, biological and/or exposure point of view (Lahr et al., 2023). So, in contrast to reliability, relevance is to a high degree determined by the goal and regulatory framework of the assessment.

# STRUCTURE OF THE CATS

Reliability – several subsets of criteria

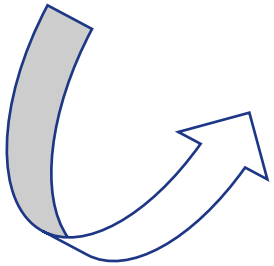
- Control performance
- Test item
- Test location
- Biological system
- Experimental system and exposure conditions
- Statistical design
- Study results


Relevance

- Study link to the GAP (exposure, conditions, location)
- Effect observed relevant for the SPG (e.g. HB colony)
- Duration allows for responding to the regulatory purpose

# THE CRITERION CATEGORIZATION AND WEIGHT

With respect to the information available in the study, different types of criteria are distinguished:



**Criterion categorization**

- Yes, fulfilled
- Yes, partially fulfilled
- No, not fulfilled

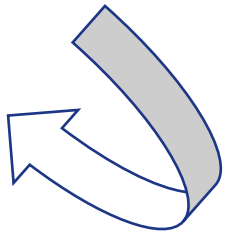
Criterion category explained in the handbook

- Not reported → Count as not fulfilled in the overall scoring
- Not applicable → To adapt to possible deviations of the study protocol/to literature studies

**Importance**

- Key criteria
- Other criteria

With respect to their importance, two different types of criteria are distinguished



# EXAMPLE OF RELIABILITY ASSESSMENT

Criterion categorization

Importance/weigh in the assessment

Reliability assessment of mesocosm studies with aquatic organisms

| Criterion code | | YES, criterion fulfilled | Yes, but criterion partly fulfilled | NO, criterion not fulfilled | Criterion not reported | Criterion not applicable | Type of criteria | Rationale |
|---|---|---|---|---|---|---|---|---|
| | **General information** | | | | | | | |
| G-1 | Is a study plan followed? Are deviations from the study plan reported and appropriately justified? | | x | | | | Other criteria | |
| | **Validity criteria and control response** | | | | | | | |
| V-1 | Are appropriate negative controls included and does the community present in the controls remain representative for a realistic freshwater community (including sensitive and/or vulnerable species) for edge-of-field surface waters during the whole test duration? | x | | | | | Key | |
| V-2 | If a positive control is included, is it appropriate and are effects demonstrated? | x | | | | | Key | |
| | **Test item** | | | | | | | |
| T-1 | Is the test substance or test item clearly identified? Is the formulation type and amount of active substance reported? | x | | | | | Key | |
| T-2 | Are test results reported for the appropriate form of the compound? | x | | | | | Key | |
| T-3 | Are test results reported for the relevant compounds (i.e. active substance(s) and/or metabolites ? | x | | | | | Key | |
| | **Test location** | | | | | | | |
| L-1 | Are the geographical location and source of environmental materials (from where water, sediments and organisms are originated, i.e., site of origin) specified, and uncontaminated? | | | x | | | Key | |
| L-2 | Are the transport conditions adequate so that stress to the test organisms caused by long-distance transportation is avoided? | | | x | | | Other criteria | |
| | **Biological system** | | | | | | | |
| O-1 | Is the biological system (e.g., species, appropriate traits, trophic level) adequately described (i.e., at the appropriate minimum taxonomic resolution) at least: i) before the application, ii) at the start of the study, iii) at different time intervals during the study, and iv) at the end of the study? | x | | | | | Key | |
| O-2 | Has the colonization and introduction of biota in the mesocosm been successful and resulted into a representative community at the end of the pre-treatment phase? | x | | | | | Key | |
| O-3 | The introduction (and potential exclusion) of biota needs to be restricted to the pre-treatment phase. Was this the ca | | x | | | | Other criteria | |
| | **Experimental system, exposure concentrations and conditions** | | | | | | | |
| E-1 | The characteristics of the test system (e.g., artificial pond/stream, dimensions, history of test system, wa agement) should not influence the test results. Was this the case? | | x | | | | Other criteria | |
| E-2 | Has the test water and test sediment been adequately described and characterized (e.g., orga content)? | | x | | | | Other criteria | |
| E-3 | Are application scheme, application mode and dose appropriately described (e.g., expo , pouring/spraying) and are all dosing solutions verified? | | x | | | | Key | |

Enter only an “X” per criterion

Notes from the evaluator – if could be helpful for supporting a different final classification

# FINAL SCORING IN EFSA CATS

## Example for reliability – overall scoring system

**Classification reliability**

| | |
|---|---|
| Class 1<br>(high reliability) | All key criteria marked as fulfilled.<br>AND<br>At least 67% of the other criteria marked as fulfilled AND none of the remaining other criteria marked as not fulfilled or not reported. |
| Class 2<br>(moderate reliability) | One or more key criteria marked as partly fulfilled and the remaining key criteria marked as fulfilled.<br>AND<br>Only 10% of the other criteria marked as not fulfilled or not reported. |
| Class 3<br>(low reliability) | All remaining cases:<br>One or more key criteria marked as not fulfilled.<br>OR<br>More than 10% of the other criteria marked as not fulfilled or not reported. |

Reference to the type of criteria by considering its importance

Reference to the relative weight of the criteria under the same category

# THE CRED METHODOLOGY

## RELIABILITY or internal validity

**Table 1.** Reliability categories[a]

| Score | Description |
|---|---|
| R1 | Reliable without restrictions: All critical reliability criteria for this study are fulfilled. The study is well designed and performed, and it does not contain flaws that affect the reliability of the study. |
| R2 | Reliable with restrictions: The study is generally well designed and performed, but some minor flaws in the documentation or setup may be present. |
| R3 | Not reliable: Not all critical reliability criteria for this study are fulfilled. The study has clear flaws in study design and/or how it was performed. |
| R4 | Not assignable: Information needed to make an assessment of the study is missing. This concerns studies that do not give sufficient experimental details and that are only listed in abstracts or secondary literature (books, reviews, etc.) or studies of which the documentation is not sufficient for assessment of reliability for one or more vital parameters. |

[a] Adapted from Klimisch et al. 3.

## RELEVANCE or external validity

**Table 3.** Relevance categories

| Score | Description |
|---|---|
| C1 | Relevant without restrictions: The study is relevant for the purpose for which it is evaluated. |
| C2 | Relevant with restrictions: The study has limited relevance for the purpose for which it is evaluated. |
| C3 | Not relevant: The study is not relevant for the purpose for which it is evaluated. |
| C4 | Not assignable: Studies that do not give sufficient details since the result is presented in abstracts or secondary literature (books, reviews, etc.) or studies of which the documentation is not sufficient for assessment of relevance for one or more vital parameters. |

# OVERALL SCORE AND OUTCOME

**Relevance criteria**

| | Key criteria | Other criteria |
|---|---|---|
| No. fulfilled | 2 | 7 |
| No. partly fulfilled | 0 | 3 |
| No. not fulfilled | 1 | 4 |
| No. not reported | 0 | 0 |
| No. not applicable | 0 | 1 |
| **TOTAL filled in the tool** | 3 | 15 |

If "Check!" appears in the adjacent cells, then one or more criterion were not classified properly. Please revise the assessment.

| | Key criteria | Other criteria |
|---|---|---|
| % fulfilled | 67% | 47% |
| % partly fulfilled | 0% | 20% |
| % not fulfilled and/or not reported | 33% | 27% |

Note that 'not reported' is counted as 'not fulfilled'

| **Overall relevance of the study** | **Class description** |
|---|---|
| | High relevance |

**Classification relevance**

| | |
|---|---|
| Class 1 (high relevance) | All key criteria marked as fulfilled. AND At least 67% of the other criteria marked as fulfilled AND none of the remaining other criteria marked as not fulfilled. |
| Class 2 (moderate relevance) | One or more key criteria marked as partly fulfilled and the remaining key criteria marked as fulfilled. AND Only 10% of the other criteria marked as not fulfilled. |
| Class 3 (low relevance) | All remaining cases: One or more key criteria marked as not fulfilled. OR More than 10% of the other criteria marked as not fulfilled. |

**Critical Appraisal Tool (CAT) - Reliability and relevance assessment of mesocosm studies with aquatic organisms**

*Adapted from Lahr et al., 2023. See also EFSA's general meeting PREV 32 (EFSA, 2025).*

**Evaluation details**

**Study information** Add full study reference
**Evaluator information** Add evaluator or institution details
**Date of evaluation** Add date of evaluation
**Study conducted under GLP?** [Select one]

The cells colored in blue and green below indicate the reliability and relevance assessments generated after the assessment tool has been completed. They will be automatically filled once the evaluation is concluded and the evaluator selects the final classification. If the risk assessor disagrees with the score provided by the tool, a robust justification must be entered in cells F15 and F18.

| | Reliability score generated by the tool | Reliability based on experts' judgement | Justification of selection of different reliability score |
|---|---|---|---|
| **Reliability assessment** | High reliability | [Select one] | *Justification to be given by the risk assessor* |

| | Relevance score generated by the tool | Relevance based on expert's judgement | Justification of selection of different relevance score |
|---|---|---|---|
| **Relevance assessment** | | | |
| [Description of the assessment goal (needed for relevance evaluation) (to be included in I64)] | High relevance | [Select one] | *Justification to be given by the risk assessor* |

Report different outcome based on experts' judgement

# ALL IN ONE

Administrative info of the study

Outcome of the evaluation

Relevance assessment

## Critical Appraisal Tool (CAT) - Reliability and relevance assessment of mesocosm studies with ...atic organisms

*Adapted from Lahr et al., 2022. See also EFSA's general meeting PREV 32 (EFSA, 2025).*

| | |
|---|---|
| **Evaluation details** | |
| **Study information** | Add full study reference |
| **Evaluator information** | Add evaluator or institution details |
| **Date of evaluation** | Add date of evaluation |
| **Study conducted under GLP?** | [Select one] |

The cells colored in blue and green below indicate the reliability and relevance assessments generated after the assessment tool has been completed. They will be automatically filled once the evaluation is completed and the evaluator selects the final classification. If the risk assessor disagrees with the score provided by the tool, a robust justification must be entered in cells F15 and F18.

| **Reliability assessment** | **Reliability score generated by the tool** | **Reliability based on experts' judgement** | **Justification of selection of different reliability score** |
|---|---|---|---|
| | High reliability | [Select one] | *Justification to be given by the risk assessor* |

| **Relevance assessment** | **Relevance score generated by the tool** | **Relevance based on expert's judgement** | **Justification of selection of different relevance score** |
|---|---|---|---|
| [Description of the assessment goal (needed for relevance evaluation) (to be included in I64)] | High relevance | [Select one] | *Justification to be given by the risk assessor* |

### General considerations/instructions

Before evaluating the test, please check the physicochemical characteristics of your compound (handbooks/general sources). What is the solubility, log KOW, pKa? Is the compound volatile? Does it hydrolyse, photolyse, etc.? What is the mode of action (MoA) of the active? What are relevant metabolites and their properties? If any properties could influence study setup or results these should be mentioned in the rationale box in row 27 (cell I27) for the general information.

Before evaluating the study, please check whether the test is performed to a guideline method or a guidance document. **This tool needs to be used with the corresponding handbook**. Mark your evaluation per criterion with an 'x' in the respective box. An explanation of the reasons for the classification should be added to the rationale box.

If the same study is used in different regulatory frameworks, the reliability assessment can stand the same; however, the relevance assessment needs to be adapted to the pertinent risk assessment situation.

### Reliability assessment of mesocosm studies with aquatic organisms

| Criterion code | | YES, criterion fulfilled | Yes, but criterion partly fulfilled | NO, criterion not fulfilled | Criterion not reported | Criterion not applicable | Type of criteria | Rationale |
|---|---|---|---|---|---|---|---|---|
| | **General information** | | | | | | | |
| G-1 | Is a study plan followed? Are deviations from the study plan reported and appropriately justified? | | x | | | | Other criteria | |
| | **Validity criteria and control response** | | | | | | | |
| V-1 | Are appropriate negative controls included and does the community present in the controls remain representative for a realistic freshwater community (including sensitive and/or vulnerable species) for edge-of-field surface waters during the whole test duration? | x | | | | | Key | |
| V-2 | If a positive control is included, is it appropriate and are effects demonstrated? | x | | | | | Key | |
| | **Test item** | | | | | | | |
| T-1 | Is the test substance or test item clearly identified? Is the formulation type and amount of active substance reported? | x | | | | | Key | |
| T-2 | Are test results reported for the appropriate form of the compound? | x | | | | | Key | |
| T-3 | Are test results reported for the relevant compounds (i.e. active substance(s) and/or metabolites ? | x | | | | | Key | |
| | **Test location** | | | | | | | |
| L-1 | Are the geographical location and source of environmental materials (from where water, sediments and organisms are originated, i.e., site of origin) specified, and uncontaminated? | | | x | | | Key | |
| L-2 | Are the transport conditions adequate so that stress to the test organisms caused by long-distance transportation is avoided? | | | x | | | Other criteria | |
| | **Biological system** | | | | | | | |
| O-1 | Is the biological system (e.g., species, appropriate traits, trophic level) adequately described (i.e., at the appropriate minimum taxonomic resolution) at least: i) before the application, ii) at the start of the study, iii) at different time intervals during the study, and iv) at the end of the study? | x | | | | | Key | |
| O-2 | Has the colonization and introduction of biota in the mesocosm been successful and resulted into a representative community at the end of the pre-treatment phase? | x | | | | | Key | |
| O-3 | The introduction (and potential exclusion) of biota needs to be restricted to the pre-treatment phase. Was this the case? | | x | | | | Other criteria | |
| | **Experimental system, exposure concentrations and conditions** | | | | | | | |
| E-1 | The characteristics of the test system (e.g., artificial pond/stream, dimensions, history of test system, water management) should not influence the test results. Was this the case? | | x | | | | Other criteria | |
| E-2 | Has the test water and test sediment been adequately described and characterized (e.g., organic matter content)? | | x | | | | Other criteria | |
| E-3 | Are application scheme, application mode and dose appropriately described (e.g., exposure pattern, pouring/spraying) and are all dosing calculations verified? | | x | | | | Key | |
| E-4 | Are methods for extraction and analysis of the test substance and transformation products, including limit of analytical detection/quantification, described and appropriate? | | | x | | | Other criteria | |
| E-5 | Have the methods used for sampling and preservation of chemical samples been adequately described and performed? | | | x | | | Other criteria | |
| E-6 | Were all exposure concentrations monitored at appropriate time intervals? | | x | | | | Other criteria | |
| E-7 | Were all exposure concentrations monitored in all relevant environmental compartments (e.g., water, sediment)? | | | | | x | Other criteria | |
| E-8 | Were the controls uncontaminated (i.e., test item < LOQ)? | x | | | | | Key | |
| E-9 | Were biota monitored at appropriate time intervals in all relevant environmental compartments (e.g., water, sediment)? | | x | | | | Other criteria | |
| E-10 | Have the methods used for observing, sampling and preserving biological samples been adequately described and performed? | x | | | | | Key | |
| E-11 | If extreme weather conditions (e.g., rain, extreme temperatures) have been recorded during the study, could they influence the results? | | x | | | | Other criteria | |
| E-12 | Are abiotic conditions monitored before and during test and within normal range (e.g., sunlight, temperature, pH, oxygen, conductivity) and similar between replicates? | | x | | | | Other criteria | |
| | **Statistical Design and Study Results** | | | | | | | |
| S-1 | Are appropriate hypothesis tests (for univariate analysis on single taxa) used? | x | | | | | Key | |
| S-2 | Is the statistical power appropriate and the MDD reported? | x | | | | | Key | |
| S-3 | Are statistical results properly elaborated? | x | | | | | Other criteria | |
| S-4 | Are raw data reported with an appropriate level of detail to check the calculation of endpoints? | x | | | | | Key | |
| S-5 | There should be no (or negligible) statistical differences in abundance/biomass in treatment and control during or at the end of the pre-treatment phase, i.e., at the start of the experiment/first application. | x | | | | | Other criteria | |

**Before evaluating the test for relevance, indicate in the box under 'rationale' (to the right) why you are evaluating this study. The relevance of endpoints from a study might be different for different purposes (e.g., PPP marketing authorization, EQS derivation, science) , also depending on the framework for which the evaluation is requested. "Regulatory risk assessment" is the default answer. Additional information can be added if desired. After the evaluation, if the study is needed for another purpose, the relevance evaluation should be repeated and reported separately.**

*[Description of the assessment goal (needed for relevance evaluation) (to be included i*

| Criterion code | | YES, criterion fulfilled | Yes, but criterion partly | NO, criterion not fulfilled | Criterion not reported | Criterion not applicable | Type of criteria | Rationale |
|---|---|---|---|---|---|---|---|---|
| R-1 | Is the experimental setup adequate for the relevant tested populations? | | x | | | | Other criteria | |
| R-2 | Are the measured variables appropriate for the investigated effects or the mode of action of the test substance? | | x | | | | Other criteria | |
| R-3 | Are the measured variables relevant on a population level (e.g. change of biomass or shoot length or abundance)? | | x | | | | Other criteria | |
| R-4 | Are effects at community level considered (that were not detectable at population/ species level) ? | x | | | | | Other criteria | |
| R-5 | Can internal recovery be separated from external recovery ? | x | | | | | Other criteria | |
| R-6 | Does the duration of the test allow to assess long-term effects and internal recovery ? | x | | | | | Other criteria | |
| R-7 | In case of a formulation, other mixture, salt or transformation product: Is the item tested representative of the one being evaluated? | x | | | | | Key | |
| R-8 | Is the tested exposure pattern relevant for the intended use of the product under evaluation? | | | x | | | Other criteria | |
| R-9 | If R-8 is fulfilled, is the concept of the envelope curve being applied and fulfilled ? | | | x | | | Key | |
| R-10 | Are the type of habitat (e.g., mesocosm/stream) and environmental compartments (water, sediment) relevant for the regulatory purpose? | | | x | | | Other criteria | |
| R-11 | Is the study representative of an EU biogeographical region, regulatory zone, or EU agricultural landscape? | | | x | | | Other criteria | |
| R-12 | Is the study suitable for the intended use, e.g., is the study period appropriate for assessing the risk of the intended use (e.g., with regard to weather conditions, presence of life stages of sensitive/vulnerable taxa)? | | | x | | | Other criteria | |
| R-13 | Are the taxonomic groups and/or species relevant for the regulatory purpose? | x | | | | | Key | |
| R-14 | If the endpoint of the mesocosm study is based upon recovery, is the risk for vulnerable species in the field like EPT (e.g., Ephemeroptera, Plecoptera, Trichoptera) species considered? | x | | | | | Other criteria | |
| R-15 | Are the measured variables (e.g. abundance, biomass) or the magnitude of effects relevant for the regulatory purpose (e.g. EC10, EC20, EC50, NOEC, effect classes)? | | | | | x | Other criteria | |
| R-16 | Are effects statistically detectable for a sufficient number of sensitive species (e.g., for deriving an effect threshold for an insecticidal MoA for at least 8 sensitive taxa from the relevant organism groups with effect class 1, 2 or 3)? | x | | | | | Other criteria | |
| R-17 | Are effects detectable at relevant time points for relevant species and taxonomic groups, also when considering the MoA of the PPP? | x | | | | | Other criteria | |
| R-18 | Can internal recovery of detected effects on relevant species be shown in a relevant period of time in order to apply an ERO-RAC? | x | | | | | Other criteria | |

### Overall assessment of the study

To update the data in the grey-shaded cells, please follow these steps: i) Go to the 'Data' tab at the top of your workbook, ii) Click on the 'Refresh All' button. This will pull the latest data into the grey-shaded areas. Remember to perform this action every time you modify any criteria to ensure that the counting is current.

**Please enter the class description of the overall reliability and relevance (i.e., low, moderate, high) of the study in the respective cells following the instructions given below. This will automatically transfer to the summary information above.**
**If there is any reason to change the classification given by the tool, please include the reasoning in the summary part at the top of this excelsheet.**

IMPORTANT FIELD **Is there any criterion flagged as not aplicable? (Yes/No)** No

#### Reliability criteria

| | Key criteria | Other criteria |
|---|---|---|
| No. fulfilled | 12 | 2 |
| No. partly fulfilled | 1 | 6 |
| No. not fulfilled | 1 | 3 |
| No. not reported | 0 | 0 |
| No. not applicable | 0 | 1 |
| **TOTAL filled in the tool** | 14 | 14 |

If "Check!" appears in the adjacent cells, then one or more criterion were not classified properly. Please revise the assessment.

| | Key criteria | Other criteria |
|---|---|---|
| % fulfilled | 86% | 14% |
| % partly fulfilled | 7% | 57% |
| % not fulfilled and/or not reported | 7% | 21% |

Note that 'not reported' is counted as 'not fulfilled'

**Overall reliability of the study** — **Class description**: High reliability

**Classification reliability**

| | |
|---|---|
| Class 1 (high reliability) | All key criteria marked as fulfilled. AND At least 67% of the other criteria marked as fulfilled AND none of the remaining other criteria marked as not fulfilled or not reported. |
| Class 2 (moderate reliability) | One or more key criteria marked as partly fulfilled and the remaining key criteria marked as fulfilled. AND Only 10% of the other criteria marked as not fulfilled or not reported. |
| Class 3 (low reliability) | All remaining cases: One or more key criteria marked as not fulfilled. OR More than 10% of the other criteria marked as not fulfilled or not reported. |

#### Relevance criteria

| | Key criteria | Other criteria |
|---|---|---|
| No. fulfilled | 2 | 7 |
| No. partly fulfilled | 0 | 3 |
| No. not fulfilled | 1 | 4 |
| No. not reported | 0 | 0 |
| No. not applicable | 0 | 1 |
| **TOTAL filled in the tool** | 3 | 15 |

If "Check!" appears in the adjacent cells, then one or more criterion were not classified properly. Please revise the assessment.

| | Key criteria | Other criteria |
|---|---|---|
| % fulfilled | 67% | 47% |
| % partly fulfilled | 0% | 20% |
| % not fulfilled and/or not reported | 33% | 27% |

Note that 'not reported' is counted as 'not fulfilled'

**Overall relevance of the study** — **Class description**: High relevance

**Classification relevance**

| | |
|---|---|
| Class 1 (high relevance) | All key criteria marked as fulfilled. AND At least 67% of the other criteria marked as fulfilled AND none of the remaining other criteria marked as not fulfilled. |
| Class 2 (moderate relevance) | One or more key criteria marked as partly fulfilled and the remaining key criteria marked as fulfilled. AND Only 10% of the other criteria marked as not fulfilled. |
| Class 3 (low relevance) | All remaining cases: One or more key criteria marked as not fulfilled. OR More than 10% of the other criteria marked as not fulfilled. |

Reliability assessment

Detailed calculation of the evaluation – automatically done

# LATEST IMPROVEMENTS OF THE EFSA CATS

- Allow for flexibility – regarding the applicability and the possibility of reporting **experts' judgement**, for each criterion and for the whole assessment
- More user-friendly tools – "all-in-one" excel sheet
- **Automatic counting** of the different sets of criterion combination
- Direct reporting of the outcome
- **Quality check** included - the tool tells you if there is a revision needed

# WHY ARE EFSA CATS RELEVANT

- The EFSA CAT handbooks are a **mature, agreed and comprehensive guideline** for the appraisal of certain non-standard studies for which OECD TG/GD are not available yet
- They are not *per se* "**checklists**" but inform on all relevant parameters to be reported and appraise in a study evaluation
- Offer a **basis for digital solutions** (e.g., web-based tools, AI-supported appraisal) to improve efficiency and consistency across assessments/assessors
- They are **living docs** i.e., to be adapted for future cases, if pertinent
- New CATs are under development – CATs in the area of ED assessment

# STAY CONNECTED

SUBSCRIBE TO
efsa.europa.eu/en/news/newsletters
efsa.europa.eu/en/rss
Careers.efsa.europa.eu – job alerts

LISTEN TO OUR PODCAST
Science on the Menu –Spotify, Apple Podcast and YouTube

FOLLOW US ON BLUESKY
@efsa.bsky.social
@efsa-animals.bsky.social @efsa-plants.bsky.social

FOLLOW US ON LINKEDIN
Linkedin.com/company/efsa

FOLLOW US ON INSTAGRAM
@one_healthenv_eu

CONTACT US
efsa.europa.eu/en/contact/askefsa

EUROPEAN FOOD SAFETY AUTHORITY