INCORPORATION OF SEDIMENT- AND SOIL-SPECIFIC ASPECTS IN THE CRITERIA FOR REPORTING AND EVALUATING ECOTOXICITY DATA (CRED)

MCarmen Casado-Martinez1, Gilda Dell'Ambrogio1,2, Sophie Campiche3, Alexandra Kroll4, Eva Lauber4, Mireia Marti-Roura4,5, Leire Mendez-Fernandez6, Mathieu Renaud1, Alena Tierbach4,5 Michel Wildi1,5, Janine WY Wong1, Inge Werner4, Marion Junghans4, Benoit JD Ferrari1,4

1 Swiss Centre for Applied Ecotoxicology. EPFL-ENAC-IIE-GE, Station 2, 1015 Lausanne, Switzerland.
2 Present address: Agroscope Reckenholz, Reckenholzstrasse 191, 8046 Zurich, Switzerland.
3 EnviBioSoil, Gollion, Switzerland.
4 Swiss Centre for Applied Ecotoxicology. Überlandstrasse 133, 8600 Dübendorf, Switzerland.
5 Present address: Federal Office for the Environment, Monbijoustrasse 40, CH-3003 Bern, Switzerland.
6 Dpt. Genetics, Physical Anthropology and Animal Physiology, University of the Basque Country (UPV/EHU), Box 644, 48080, Bilbao, Spain.

* Corresponding author: carmen.casado@centreecotox.ch

Acknowledgements:
The authors would like to thank Muris Korkaric and Robert Pasanen-Kase for their comments on a previous version of this manuscript. There are no funders to report.

Data availability statement:
All data are included as Supplementary Information.

This document is the accepted manuscript version of the following article:
INTEGRATION OF SEDIMENT- AND SOIL-SPECIFIC ASPECTS IN THE CRITERIA FOR REPORTING AND EVALUATING ECOTOXICITY DATA (CRED)

Abstract: In environmental risk assessment either for registration purposes or for retrospective assessments of monitoring data, the hazard assessment is predominantly based on effect data from ecotoxicity studies. Most regulatory frameworks require studies used for risk assessment to be evaluated for reliability and relevance. Historically, the Klimisch methodology was used in many regulatory procedures where reliability needed to be evaluated. More recently, the Criteria for Reporting and Evaluating Ecotoxicity Data (CRED) have been developed for aquatic ecotoxicity studies, providing more detailed guidance on the evaluation and reporting of not only the reliability but also the relevance of a scientific study. Here, we discuss the application of the CRED methodology for assessing sediment and soil ecotoxicity studies, addressing important sediment- and soil-specific criteria that should be included as part of the CRED evaluation system. We also provide detailed recommendations for the design and reporting of sediment and soil toxicity studies that can be used by scientists and researchers wishing to contribute ecotoxicological data for effect assessments carried out within regulatory frameworks.

KEYWORDS: Relevance evaluation; Reliability evaluation; Sediment ecotoxicology; Soil ecotoxicology
INTRODUCTION

Ecotoxicological data are a fundamental component of environmental risk assessments (ERA). While requirements may differ between regulatory frameworks (e.g., European Commission [EC], 2018; EC SANCO, 2002; Marti-Roura et al., 2023), the general ERA paradigm is similar insofar as predicted or measured environmental concentrations (P/MEC) are compared to a hazard assessment derived from ecotoxicological data. Although less commonly than water, sediment and soil are part of generic predictive ERAs for chemical registrations (Deneer et al., 2013; EC SANCO, 2002; European Chemical Agency [ECHA], 2008, 2014, 2015, 2017, 2023; European Food Safety Authority [EFSA], 2015, 2017; European Medicines Agency Committee for Medicinal Products for Human Use, 2016, 2018). Ecotoxicological data are also used to set protection values for monitoring and site characterization in retrospective risk assessment of soils and sediments (Canadian Council of Ministers for the Environment [CCME], 2006; EC, 2018; NEPC, 2013; RIVM, 2007; USEPA, 2005). The requirements for effect assessment differ between environmental compartments and, for the same environmental compartment, may differ between regulatory frameworks.

In most regulatory frameworks, available ecotoxicological data must undergo an evaluation of reliability and relevance. Until recently, the methodology developed by Klimisch et al. (1997) was used to evaluate reliability of data from both toxicological as well as ecotoxicological studies (EC, 2018; ECHA, 2011, 2023; European Medicines Agency, 2016; RIVM, 2007). The Klimisch score assigns studies or data to one of four categories. “Reliable without restrictions” is assigned when data are generated according to generally valid and/or internationally accepted testing guidelines or comparable methods. “Reliable with restrictions” is assigned when the test does not totally comply with the specific testing guideline but is well documented and
scientifically acceptable. “Not reliable” are data generated from studies in which there are interferences in the exposure conditions or an unacceptable method is used. “Not assignable” is attributed to studies which lack the detailed information needed to evaluate reliability. Data categorized as “reliable without restrictions” and “reliable with restrictions” are generally considered adequate for use in environmental hazard and risk assessments, while those categorized as “not reliable” and “not assignable” are not accepted for regulatory use but may be considered as supporting information.

Other evaluation methods based on the Klimisch method have since been developed to improve its limited guidance, and to decrease dependency of the evaluation on expert judgement (see reviews by Moermond et al., 2017; Roth and Ciffroy 2016). An alternative method was developed by Moermond et al. (2016), the Criteria for Reporting and Evaluating Ecotoxicity Data (CRED) methodology, which focuses on improving the evaluation of reliability and relevance assessments, specifically for aquatic toxicity studies, by increasing reproducibility, consistency and transparency of evaluation results. In addition to establishing clear evaluation criteria, the CRED methodology is accompanied by an extensive guidance document for risk assessors on how to apply the criteria. The CRED criteria can also be used as a checklist by those responsible for conducting a study to help ensure accurate and complete reporting of study results. In this way, the usability of the peer-reviewed literature for regulatory purposes can be improved.

As stated by Moermond et al. (2016), CRED addressed ecotoxicity studies in the water compartment, but could be adapted to other compartments such as soil and sediments. While a number of original criteria can be easily applied to soil and sediment data, others need additional clarification or adaptation and additional criteria are required for sediment- and soil-specific issues. This paper intends to fill this gap.
A checklist of reporting recommendations is provided as supporting information for researchers to use when designing and publishing their sediment and soil ecotoxicity studies, to ensure that risk assessors can make the best use of their results. This is particularly relevant for soil and sediment, given the limited number of standard test guidelines currently available (especially for sediment tests) and the incomplete reporting of sediment and soil toxicity studies in the scientific literature. Studies published in the scientific literature often report data for endpoints and exposure conditions that cannot be categorized as relevant and reliable with/without restrictions categories, which means they cannot be used for regulatory purposes. As a result, a limited dataset on the toxicity of substances to soil and sediment organisms is available, which hampers effect assessments and consequently ERAs for these environmental compartments.

**CRITERIA FOR ASSESSING RELEVANCE AND RELIABILITY OF ECOTOXICOLOGICAL STUDIES WITH SEDIMENTS AND SOILS**

The original CRED evaluation method for aquatic ecotoxicity studies (Moermond et al., 2016) was adapted for sediment and soil ecotoxicity data building on ECHA, EFSA, and other documents from national and international competent authorities, validated protocols and standard operating procedures for sediment and soil toxicity testing from Environment Canada, ASTM, OECD and ISO (e.g., CCME, 2006; ECHA, 2014, 2023; EFSA, 2015; NEPC, 2013; RIVM, 2007; USEPA, 2000, 2011), and additional information from the scientific literature (e.g., Beketov et al., 2013; Breton et al., 2009; Diepens et al., 2014). For a detailed description of general criteria common to all three environmental compartments the reader is referred to the original CRED by Moermond et al. (2016). Here, we provide supportive information
and clarifications on how to apply the CRED general criteria and provide additional criteria intended for the evaluation of sediment and soil ecotoxicity studies. Excel files are provided as Supplementary Information, Appendix A for sediment and Appendix B for soil, together with reporting recommendations. The approach is suitable for assessing sediment and soil ecotoxicity studies performed with sediment and in-soil micro-, meso- and macrofauna and plants. For studies with microorganisms, expert judgment and case-by-case decision-making is needed throughout the relevance and in particular the reliability assessments, although the proposed approach can be used to guide the assessor through the study evaluation. Note that although we refer to “study reliability” and “study relevance” for simplification purposes, these categories relate to the reliability and relevance of an effect datum and not the entire study (one study may report a reliable as well as an unreliable endpoint and a relevant as well as not relevant endpoint).

**Relevance evaluation of sediment and soil toxicity studies**

Four relevance categories are presented in the original CRED (Moermond et al., 2016) (Table 1): relevant without restrictions (C1), relevant with restrictions (C2), not relevant (C3) and not assignable (C4). In order to assess which of the four categories a study belongs to, a total of 12 criteria are proposed for sediment and soil ecotoxicity studies (Table 2) (Moermond et al., 2016). Criteria can relate to biological relevance (Criterion C#1 to Criterion C#9) or exposure relevance (Criterion C#10 to C#12). They should help the risk assessor to answer the key question of whether it is justified to use a particular ecotoxicological study directly or indirectly in the context of an ERA, taking into account the specific regulatory protection goal mandated by legislation, regulation or policy (Rudén et al., 2017). Most regulatory frameworks include separate ERA schemes for different environmental compartments (aquatic,
sediment, soil). Thus, ecotoxicological data should be assessed according to specific relevance criteria for the sediment and soil compartment.

**General information.** An initial step for assessing the reliability and relevance of ecotoxicity studies is to document the physico-chemical properties of the tested compound, including solubility, log \( K_{\text{OW}} \), log \( K_{\text{OC}} \), pKa, potential hydrolysis, photolysis, etc. This information is needed to assess compliance with reliability and relevance criteria and, if not reported with the study results, should be gathered by the assessor from suitable sources. It is recommended that data on physico-chemical properties be also reviewed for relevance and reliability according to the intended use.

**Biological relevance.** In general, only species that act as ecological representatives for a given environmental compartment are acceptable as test organisms for risk assessment (Criterion C#1; ECHA, 2023). A review of sediment ecotoxicological test methods is available in Diepens et al. (2014), Simpson and Batley (2016) and Leppanen et al. (2024). An overview of soil ecotoxicological test methods is available in van Gestel (2012) and standard methods are provided in Römbke and Martin-Laurent (2020). In general, species used or recommended in guideline methods for testing a specific environmental compartment are considered relevant *per se*. For non-standard toxicity test species and methods their relevance has to be evaluated using expert judgement. Species that are similar to, or share characteristics with, either standard test species or taxonomically related field species are considered relevant. Species-specific life traits and natural history should be considered. Species are more relevant to a particular environmental compartment if they spend most or a substantial or critical part of their lifecycle in contact with that environmental compartment. For example, midge larvae, sediment-burrowing invertebrates or rooted macrophytes are preferred for sediment risk assessment over species that may be less exposed to
contaminated sediments (e.g., *Daphnia magna* or *Cerodaphnia dubia*). However, an organism's relevance may differ between environmental compartments. For example, microorganisms are much less considered in sediment effect assessments (Diepens et al., 2014) in comparison to soil effect assessments. Given the important ecological functions that microorganisms perform in soil and sediment, microbial ecotoxicology is developing rapidly and warrants further consideration in impact assessments in the near future (Pesce et al. 2020; Karpouzas et al. 2022).

In addition to being relevant for the environmental compartment at issue, biological relevance should address whether the test species is appropriate for the tested compound (Criterion C#2). For both soil and sediment, specific recommendations or requirements for the selection of test organisms for effect assessment are provided in regulatory and standard guidance documents. For prospective risk assessment in Europe, tiered decision schemes are available for the required test species for soil or selection of relevant species for sediment (EC SANCO, 2002; EFSA, 2015). Specific guidance for the selection of terrestrial organisms for effect assessment can be found in ECHA (2023) and for the ecotoxicological characterization of soils and soil materials in International Organization for Standardization (ISO) (2019). Recommendations also relate to how substances interact with the environmental matrix, where for strongly adsorbing or binding substances species that feed on particles (e.g., oligochaetes, earthworms) are preferred (ECHA, 2023). In general, information on less sensitive species (if available) may also become relevant, e.g., for the implementation of species sensitivity distributions. However, the different sensitivities of soil and sediment organism groups to specific classes of substances are not always clearly defined.
The relevance of the tested life-stage is assessed in Criterion C#3. If only part of the lifecycle takes part in the soil or sediment, the appropriate life-stage that lives in contact with the environmental compartment at issue should be tested. Sensitivity may also differ substantially during different life-stages of the same species. Early life-stages are usually preferred for standard tests over later life-stages and they are more sensitive than adults, although exceptions may occur (Collyard et al., 1994). Juveniles or (sub)adults are usually preferred for reproduction tests, e.g., with oligochaetes and soil invertebrates. In case adaptations from the standard guideline are made, expert judgement is required to evaluate if the results are still relevant.

The relevance of the endpoint should also be carefully evaluated (Criterion C#4 to Criterion C#7). The reported endpoint should be considered appropriate for the specific regulatory purpose and protection goal. Survival, growth and reproduction are relevant endpoints if standard test species are used. They are assumed to account for biomass or abundance when population is the protected ecological entity (Rudén et al., 2017). Bioaccumulation may become relevant but only if the objective is protection against secondary poisoning. It may however be used as an additional line of evidence for substance uptake in test organisms. Non-standard endpoints may be accepted, but their ecological relevance must be carefully assessed. If non-standard endpoints are very different from the standard endpoints, then these must be scientifically justified. Standardized behavioral observations such as avoidance for earthworms (ISO, 2008) and collembola (ISO, 2011) are considered particularly relevant for soil and are accepted as predictors of effects at the population level (EFSA, 2017). Behavioral endpoints have not yet been standardized for sediment toxicity testing (e.g., Johns et al., 2023, Thit et al., 2020) and in principle are not considered relevant. However, sediment avoidance and lack of burrowing activity also provide indication of toxic
effects in benthic invertebrates (Amiard-Triquet, 2009). As non-standardized behavioral observations, they should not be interpreted in isolation and should be assessed using expert judgement (ECHA, 2023). Specific guidance for the use of behavioral endpoints in regulation is available from Ågerstrand et al. (2020). Regarding the use of endpoints at the cellular and molecular level including biomarker responses such as vitellogenin concentrations or gene expression, such endpoints are in principle not considered relevant and are not yet well accepted for use in the derivation of regulatory threshold values. But they can be used when a definite correlation or causal relationship with population sustainability is established; otherwise they may be used as supporting information (EC, 2018). They can also be relevant if scientific evidence clearly indicates a link with the specific mode of action of the test substance, e.g., sex ratio as indicative of endocrine-disrupting chemicals. Biomarkers might also be relevant for soil assessments when standard tests do not show effects within the standard investigated test period but, due to the mode of action of the substance, delayed or long-term effects can be expected (EFSA, 2017). For soil, exo-enzymes produced by soil microorganisms can be used as biomarkers of soil quality and are an integral part of the soil ecological balance (e.g., NEPC, 2013; RIVM, 2007). As such, microbially mediated enzymatic activities could be included in the assessment as “relevant with restrictions”, however, as fundamentally different (processes involving multiple species), these endpoints should not be merged with single species data (ECHA, 2023).

The relevance of the endpoint should also be assessed according to exposure duration and test species (Criterion C#6). Standard ecotoxicity tests have defined durations but if the duration of the exposure is different from that in the corresponding guidelines, a scientific justification should be provided or the study should not be
scored as relevant without restrictions. For non-standard test methodologies, it is important to ensure that the duration of exposure in the test is adequate for the test substance to have an effect on the test organisms. In chronic tests, the duration should cover a considerable part of the lifecycle. Especially for strongly adsorbing substances, it may take some time to reach equilibrium between the soil/sediment concentration in the test system and in the test organisms. For sediment, the relevance of short-term tests (e.g., 10 days) with benthic invertebrates (e.g., ASTM E1706-20) differs depending on the regulatory framework and endpoint considered. For example, short-term tests (e.g., *Heterocypris incongruens* 6 days ISO, 2012) should not be used solely for the derivation of sediment PNECs (ECHA, 2023); while specific recommendations are included in the guidance document for the derivation of sediment EQS based on acute effect data by increasing the assessment factor (EC, 2018). For soils, acute tests may be accepted (with a higher assessment factor) and are still many times required for earthworms, but preference is increasingly given to longer chronic exposures (CCME, 2006; EC, 2003; EC SANCO, 2002; USEPA, 2005).

The statistical significance of the magnitude of the effect and its biological relevance for the regulatory purpose is assessed in Criterion C#8. Several statistical estimates are used to express the results of ecotoxicity tests (e.g., no observed effect concentration [NOEC], EC10, EC20, EC50) but preference is given to certain estimates under certain regulatory contexts. For example, no or low effect concentrations (EC10, NOECs) are preferred for derivation of EQS or safe thresholds for long-term protection goals (e.g., EC, 2033, 2018; Marti-Roura et al., 2023; RIVM, 2007) but other estimates such as L(E)C25-50 can be used to derive more permissible effect levels (CCME, 2006; International Council of Mining and Metals [ICCM], 2016; NEPC, 2013). Historically, the most reported effect concentrations for sediment and
soil toxicity studies were NOECs derived from hypothesis-testing methods, although EC10 and EC20 are becoming more common in the scientific literature and are gaining more weight in regulation compared to NOECs according to its independence from test design (Iwasaki et al., 2015). If raw data are reported, the assessor should be able to calculate additional – statistically significant – effect concentrations to use instead of reported NOECs.

A final criterion for assessing the biological relevance of the study is related to the relevance of the experimental conditions for the tested species (Criterion C#9). In general, experiments that have been performed according to standard test guidelines and fulfil the specified test validity criteria (see reliability Criterion R#3) are considered relevant. In the case of modified experimental conditions, the relevance of exposure conditions should be further evaluated together with relevance Criterion C#10 to C#12 (see below).

*Exposure relevance.* The morphology, physiology and behavior of soil and sediment organisms determine their exposure pathways. Thus, the information gathered for assessing biological relevance of the species will also be useful in assessing whether the experimental conditions account for environmentally relevant exposure scenarios (Criterion C#10). In general, contaminated soil/sediment is the preferred route of exposure for many benthic and soil species (EFSA, 2015, 2017). Alternative exposure scenarios such as filter paper or agar plates for micro-organisms are most often not considered relevant (e.g., Marti-Roura et al., 2023; RIVM, 2007; USEPA, 2005). For some benthic invertebrates, such as bivalves, relying on large quantities of suspended matter as food (Bejarano et al., 2003), suspended matter may be a more relevant exposure scenario but results from spiked-sediment toxicity tests can still be considered relevant. However, if the route of administration is exclusively
via contaminated food, the study could be considered as supplementary information but not used in effect assessments. The addition of uncontaminated food items during exposure is common practice in both standard and non-standard toxicity tests, although the relevance of exposure conditions when test organisms are fed in long-term tests on naive (non-contaminated) food has been questioned (ECHA, 2023). Testing strategies that account for contaminated food exposure in sediment (and likely in soil) can evaluate the relative importance of food through pre-testing. Therefore, if relevant, food items should be equilibrated with contaminated sediment/soil in order to ensure that the test organisms are exposed to environmentally relevant conditions (ECHA, 2023; Natal-da-Luz et al., 2019; OECD, 2023). Standard operating procedures for such type of test are not yet available. Sediment spiking and equilibration procedures may also render study results non-relevant (see reliability criteria related to exposure conditions and specifically Criterion R#15).

The test substrate should also be appropriate for the test organisms, which can be assessed from the results of the negative toxicity controls and the test validity criteria (see reliability Criteria related to test setup R#3 to R#5). Standard test guidelines include recommendations for the selection of adequate formulation for artificial substrates for the test species (e.g., ISO, 2013a, 2013b; OECD, 2014, 2016a, 2016b, 2023). ASTM, USEPA and OSPARCOM guidelines, which prefer natural sediments for testing, include the acceptability range of sediment properties for different test species (ASTM International, 2020; OSPAR Commission, 2006; USEPA 2000). When non-standard test organisms are used, available information of the species and expert judgement are needed to evaluate whether the substrate is appropriate for the test organism.
For the registration of chemicals, artificial soil is the most commonly used substrate containing 10% peat as source of organic matter. However, this has been criticized as not being representative of most soils. Thus, lower proportions, usually 5%, of peat are accepted and already recommended in standard guidelines (OECD 2016a, 2016b). However, this is still considered as the upper end of possible organic matter content of agricultural soils and thus a possible bias of underestimating the toxicity of pesticides to soil macro- and microorganisms. Therefore, where standard artificial soil is used, a correction factor of 2 has been applied, for example in plant protection products (PPPs) regulation, for substances with log $K_{ow}$ >2, unless it is demonstrated that toxicity is independent from carbon content (EC SANCO, 2002). For sediments, test data in which availability is maximized are in general preferred, thus the assessor should evaluate whether the test sediments provide a worst-case scenario for the tested substance or not (OECD 2016a). Artificial sediment most commonly has 2% total organic carbon (TOC) and low acid volatile sulfides (AVS) concentrations. For metals, tests under aerobic conditions with low AVS levels (<1.0 µmol AVS/g d.w.) are preferred. Thus, studies performed with artificial sediments are relevant for representing worst-case scenarios for risk assessment. When natural sediments are used, AVS concentrations should be reported. Otherwise, expert judgement may be needed to evaluate if a study was conducted under oxic conditions, in which water column oxygen levels are controlled and reported.

Likewise, the exposure scenario (duration, concentrations, substance application, route of administration, exposure schedule) should also be relevant for the test substance, taking into consideration the specific properties of the substance and the way it is used (Criterion C#11). For example, OECD guidelines are available for tests with *Chironomus riparius* in spiked sediment exposure (OECD, 2023a) and in
spiked overlying waters (OECD, 2023b). However, application through overlying water is only relevant for certain regulatory purposes and when exposure is expected to occur mainly through water according to substance application and its environmental fate. For soil organisms, exposure via pore water may be relevant in certain cases, especially for compounds with high solubility (EFSA, 2017). For metals, control conditions should be relevant for natural exposure, including background concentrations (ICCM, 2016).

The relevance of the test substance should also be assessed, in particular if a formulation, mixture or salt is used for testing (Criterion C#12). If the test substance is known to degrade in the environment, the presence or absence of degradation products in the test study should also reflect relevant exposure scenarios. For sediment risk assessment, the same triggers used for the parent compound should be applied for degradation products as for the active ingredient, e.g., $\log K_{OW} > 3$.

**Reliability evaluation of sediment and soil toxicity studies**

Four reliability categories are presented in agreement with Klimisch et al. (1997) and Moermond et al. (2016) (Table 3): reliable without restrictions (R1), reliable with restrictions (R2), not reliable (R3) and not assignable (R4). R1 is attributed to data from studies that are well designed and performed, do not contain flaws that affect the reliability of the study, and full reporting is available. If the study has minor flaws, it is assigned to category R2 as reliable with restrictions. The score R3, not reliable, is assigned to studies that use methods that are not appropriate to address the research question or contain major flaws that affect the reliability of the study. Studies are classified as R4 when insufficient experimental details are provided, they are only published as abstracts, they can only be found in secondary literature or when
documentation is not sufficient for the assessment of reliability of one or more of critical parameters.

Reliability criteria proposed for the classification of studies into reliability categories relate to (Table 4): 1) test setup (Criterion R#1 to Criterion R#5), 2) test compound (Criterion R#6 to Criterion R#9), 3) test organism (Criterion R#10 and Criterion R#11), 4) exposure conditions (Criterion R#12 to Criterion R#19) and 5) statistical design and biological response (Criterion R#20 to Criterion R#23). In total 23 criteria are provided for reliability assessment. Sediment- and soil-specific criteria relate mainly to exposure characterization, to the type of solid matrix (sediment/soil) used for testing, its main characteristics, and the spiking, equilibration and aging if any prior to testing.

**Reliability criteria related to test setup.** Reliability criteria related to test setup address whether a standard or a modified standard is used (Criterion R#1) and whether the test is performed under good laboratory practice (GLP) principles (Criterion R#2). These are informative rather than critical criteria because the use of a guideline method or GLP principles do not necessarily mean that the study is reliable. A quality check should always be carried out for flaws in the design, conduct, and/or statistical interpretation. Likewise, modified standards and non-standard methods may be as reliable as studies following standard methods. Non-standard methods and modifications from standard tests have to be carefully evaluated by expert judgement as they may be an important contribution to the effect assessment for sediment and soil compartments.

If a standard method is used, clear guidance is provided on the type of controls that should be run and validity criteria that should be met to ensure study results are well-founded and deviations in study conditions during test implementation do not
compromise reliability (Criteria R#3 and R#4). Validity criteria typically refer to responses in test endpoints for negative toxicity controls (maximum mortality, minimum emergence, minimum number of juveniles, maximum coefficient of variation, etc.). Additional non-standard effects observed in the controls, like behavioral changes (e.g., lethargy, absence of feeding activity, substrate avoidance), morphological alterations (e.g., change of color or size), or any signs of harm or damage to the organisms (e.g., leaf necrosis in plants) are signs of stress and/or poor health during exposure duration, which may compromise the reliability of data (Criterion R#5). In the case of sediment ecotoxicity studies, additional validity criteria are set for the water quality parameters, such as minimum dissolved oxygen, pH, and temperature. In addition, ammonia concentrations, which may accumulate in sediment-based systems that are static for long periods, should be monitored for highly sensitive species (ECHA, 2023). Ideally, these parameters should be measured close to the sediment surface where organisms are exposed. If validity criteria are not met, the reliability of the study may be critically compromised and it should not be classified as R1. However, not all validity criteria have the same level of criticality and expert judgement can be used to assess when small deviations in, e.g., water quality parameters or non-conforming results in a control replicate do not compromise the reliability of the whole study. For non-standard tests and tests performed according to a standard guideline but using a non-standard test species, expert judgement is needed to decide if controls are within the normal range of the species.

Negative toxicity controls should be run with the same batch of test organisms and under identical exposure conditions as the treatments. If a solvent is needed for the preparation of the test concentrations, a solvent control is also required. The response in the solvent controls should ideally not differ from negative controls but if
the response differs significantly, statistics should be based on solvent control response. If this is not the case, a re-assessment of the results may still be possible for well-reported studies. A study is reliable without restrictions if appropriate controls are performed. If solvent controls are not performed a study can be classified as relevant with restrictions and still be acceptable for regulatory use based on existing information on the sensitivity of the test organism under similar solvent control conditions. If this information is not available or solvent control conditions do not correspond to test treatments, the study should be considered not assignable or not reliable.

**Reliability criteria related to test compound.** Four criteria relate to the test compound, addressing the identification and purity of the test substance and the appropriate consideration of its physico-chemical characteristics during the study. In general, a study cannot be classified as reliable without restrictions if the test substance is not clearly identified (e.g., CAS number, IUPAC name or IUPAC International Chemical Identifier; Criterion R#6) and its purity and/or origin are unknown (Criterion R#7). There are no threshold levels for purity under OECD ("suitable purity"), although the purity of the substance should be ideally >90%, otherwise the nominal test results should be corrected for purity (e.g., EC, 2018; RIVM, 2007). When effect concentrations are based on measured rather than nominal concentrations this criterion is less important, but this is rarely the case for soil studies. Expert judgement is required to assess whether the impurities may have an adverse or beneficial effect on the test organism and thus influence the outcome of the study (Criterion R#8). When a formulation is used for testing a specific compound, ideally full information for all other constituents and effect data should be considered for the sum of the constituents. However, information on components of the formulation is usually
confidential and is not available for the majority of PPPs. In case toxic effects of co-formulants or additives are known, this information should be provided by the study authors. The units of concentration should always be reported as the amount of active substance (a.s.), not the amount of product, or it should be possible to re-calculate the units into amount of a.s. based on available data. If sufficient information is provided, the use of formulations does not influence the reliability assessment of the study.

Only studies that demonstrate that substance properties have been taken into consideration for study design (Criterion R#9) are reliable without restriction (R1). If any physico-chemical properties of the substance may affect its behavior during the study (e.g., solubility, volatility, stability hydrolysis, photolysis, degradation, log \textit{K}_{\text{ow}}, adsorption), this should be reported and discussed by the authors. Standard tests should follow guidance documents as closely as possible but some flexibility may be required to adapt the study design and the experimental system to the physico-chemical properties of the chemical. Guidance on critical substance characteristics and recommendations for refinements of standard test procedures as well as results interpretation can be found in ECHA (2023). Also, standard methods for soil organisms typically include a disclaimer stating which physico-chemical properties make compounds ineligible for testing (e.g., OECD, 2016a). Results for substances that contain many constituents also present interpretation problems if the different components have different physico-chemical properties leading to heterogeneous dosing.

\textbf{Reliability criteria related to test organisms.} For a study to be reliable, the test organisms should be clearly identified and described (Criterion R#10). The name and age, life-stage, size or weight of the test organisms should be reported at a minimum. Organisms should be identified to the species level and expert judgment is required to
assess the reliability of the study if only a higher level of the taxonomic hierarchy is reported. For tests with microorganisms (e.g., bacteria and fungi communities in soil) identification to the species level is not a requirement because it is often not feasible. For field-collected organisms, the method used for taxonomic determination (physical morphology, molecular determination) should ideally be reported. For non-standard tests, body length or weight may be used for determining the age or life-stage of the test organisms if not otherwise described. Other information may be necessary in a reliability assessment depending on the test endpoints, e.g., the sex and/or reproductive stage of the test organisms when the endpoint is reproductive output. For a study to be reliable without restriction, the source of test species should be known and trusted (Criterion R#11). For microorganisms and microbial processes, the source of the test soil, its storage conditions and any incubation and inoculation procedures must be reported. Standard test organisms are most often harvested from laboratory nurseries maintained under well-established conditions, which usually ensure a good performance and comparability between test results. Nevertheless, laboratories should test the sensitivity of the test organisms on a regular basis. For field-collected organisms, common in non-standard tests, it is important to know the place of origin along with a full site characterization and details of any potential exposure of the organisms to toxic or non-toxic stress that may result in community adaptation and therefore bias in results (ICCM, 2016; Rainbow, 2002). If the test organisms are not bred in the same sediment/soil type or kept under the conditions that are used for the final test (including feeding) or they are from field populations, they should be acclimatized (e.g., at least 24 hours; OECD, 2016a; ASTM 2022) to the test conditions and this information reported (Criterion #11).
**Reliability criteria related to exposure conditions.** Reliability criteria related to exposure conditions address the appropriateness of the experimental system for the test substance and for the test species. The suitability of the test system for the test substance, taking into consideration its physico-chemical characteristics, is assessed in reliability Criterion R#12 together with Criterion R#7. For a study to be reliable, the experimental system and the experimental conditions (feeding regime, temperature, light/dark conditions, pH) should be adequate for the test species that can be controlled through validity criteria (see also Criterion R#3) and should be stable during the test (Criterion R#13). The duration of the test should also be clearly defined and adequate for the test species and the endpoint, otherwise the study cannot be considered reliable (Criterion R#17).

The results of toxicity testing with soil and sediment are also highly dependent on the solid matrix selected for the test, thus it is necessary that all parameters that may have an influence on bioavailability and therefore toxicity are characterized and well reported (Criterion R#14). For sediments, in general, the main parameters driving the bioavailability of substances in solid matrices are clay and TOC or organic matter (OM) content, Cation Exchange Capacity (CEC) and pH. At least grain size distribution and TOC or OM should be reported to classify a sediment study as reliable without restrictions. Iron and manganese oxides, AVS and simultaneously extracted metals (SEM) facilitate bioavailability corrections for metals if necessary (ICCM, 2016). In addition, ammonia in pore water, nitrogen content and percent water content are reporting requirements for certain standard toxicity tests. For soils, pH should be reported (ISO, 2014; OECD, 2016a). Also, the maximum water holding capacity (WHC) must be reported and corrected to the appropriate range for testing (usually between 40 – 60% WHC) as described in standard methods (e.g., OECD 2016a). Normalization
of toxicity data to TOC or OM may be required because they usually provide the main binding phase for many types of substances (EC, 2003, 2018; OECD, 2023a).

Spiking and equilibration may also have a great impact on the results of sediment and soil toxicity tests. A study can only be classified as reliable without restrictions if the spiking procedure is reported and appropriate to ensure reliable exposure conditions during the test duration (Criterion R#15). There is no universal protocol for sediment or soil spiking that ensures reliability of a toxicity study but standard toxicity test guidelines include some recommendations that can be used for study design and reliability assessments (e.g., OECD, 2023a for sediment, OEC, 2016a for soil). The spiking procedure should ensure proper homogenization of the test substance and therefore needs to be adapted according to the solubility of the test item (soluble in water, soluble in organic solvents or not soluble at all). If solvents are necessary, only volatile solvents should be used and — after application to the substrate — left to evaporate ensuring no dissipation of the test compound during that time. Appropriate solvent controls should be run according to standard guidelines. Most recommendations involve spiking of a small quantity of substrate (or quartz sand for artificial substrates), and then mixing it with the rest of the substrate thoroughly at the relevant exposure concentrations providing a homogenous distribution of the substrate. For spiking natural sediments with organic compounds, drying part of the sediment (e.g., 10%) is necessary before adding the test substance and then mixing it with the wet sediment is recommended (León-Paumen et al., 2008). For sediments the adequate equilibration between water and sediment must be ensured and information on equilibration time and equilibration conditions should be reported. Laboratory exposures with spiked sediments with short equilibration times are subject to high bioavailability due to loosely bound test substance and exposure through the dissolved
phase. Degradation/hydrolysis may also occur during equilibration thus a balance between equilibration and degradation should be established. Testing should ideally initiate when overlying water concentrations are stable, thus spiked sediment test design requires equilibration and exposure concentrations measured in pore water, sediment, and overlying water. Assessors should consider if a short equilibration period may have critically influenced test results when no analytical measurements of the overlying water are provided. Equilibration conditions should ideally take place under the temperature and aeration conditions used during the exposure phase. However, alternative conditions may be more adequate for certain substances, for example aeration is not recommended for substances with a Henry's law constant of 1-10 Pa ·m$^3$/mol to avoid volatilization (ECHA, 2023).

The selection of exposure concentrations is also critical to ensure reliability of ecotoxicity results, and both the range of exposure concentrations and the spacing should be appropriate (Criterion R#16). Reliability of test results is reduced when spacing between test concentrations is too large, especially when deriving a NOEC. Test concentrations should typically follow a geometric series, with spacing factors being dependent on the test organism and the experimental design. For standard guidelines specific recommendations are provided on adequate concentration distributions for different test designs and are affected by species, statistical method and steepness of the dose-response curve (e.g., OECD, 2014a, 2014b, 2023a, 2016a). For non-standard tests, expert judgment is required to evaluate the adequacy of the spacing factor but for the calculation of effect concentrations these should be determined by interpolation rather than extrapolation.

Analytical verification of tested concentrations over the duration of the study is most often required for the consideration of a sediment study as reliable without
restrictions and effect concentrations based on nominal concentrations are only acceptable if stability of exposure concentrations during exposure duration is ensured, for example from previous tests, or if supported with chemical analyses (Criterion R#18). Nominal concentrations can still be used to derive effect concentrations if measured concentrations do not vary by more than 20% between the start and end of the test. Otherwise, effect concentrations should be analytically verified following appropriate sampling strategies and, if relevant, expressed as time-weighted average concentrations or geometric mean of measured concentrations during exposure. For metals and inorganic substances, chemical analyses in overlying water are required, in particular when semi-static and static sediment toxicity tests are performed. The most recent update of recommendations for analytical verification of exposure concentrations in sediment toxicity testing for regulatory use is available in OECD 2023a. For soil ecotoxicity tests, while preferred, analytical verification of nominal concentrations is usually not a requirement, although for volatile, unstable or readily biodegradable substances, or where there is otherwise uncertainty in maintaining the nominal soil concentration, analytical measurements of the exposure concentrations at the beginning, and at the end of the test should be considered (OECD 2016a). Whenever concentrations are measured by chemical analysis, the analytical methods used should be properly described, including limit of detection and quantification. It should also describe in detail the extraction procedure, i.e., if total or soluble extractions were performed for soil and in the case of sediment which fraction was analyzed, ideally with relevant quality evaluation (Loos 2012).

Finally, the biomass loading or the number of the organisms in the test system should be within the appropriate range (Criterion #19). Biomass loading should not exceed the amount/limit for appropriate growth/reproduction of the test organisms in
the test system (e.g., shading might influence plant growth if density is too high). Additionally, the stocking density of test organisms should be low enough and the volume of the test system high enough so that the concentration of the test substance is maintained throughout the study. To avoid the potential for chemical uptake by organisms to deplete contaminant concentrations in test sediments, limits on the mass of organisms that can be added are included in test protocols, in particular for bioaccumulation tests (e.g., a loading rate of 1:10 proposed in the revised guidance ASTM (2019). This information can be used as a guideline to assess the reliability of exposure conditions. For sediments with extremely low organic carbon content where it is impractical to meet the 1:10 rate, the study report should note the exceedance of the loading rate and the additional uncertainty this may create. For standard tests with soils, the number of individuals used in each test should be according to the guidance documents and expert judgement should be used for assessing the biomass loading in case of deviation or for non-standard tests.

Reliability criteria related to statistical design and biological response. Studies are reliable when a sufficient number of replicates and organisms per replicate are used for all controls and test concentrations (Criterion R#20). For standard tests, the guideline requirements should be followed although there is some space for adjustment according to statistical considerations. When a non-guideline study is evaluated, expert judgement is needed to assess whether the study design is appropriate to obtain statistically reliable results. Also, appropriate statistical methods should be used for the determination of effect concentrations (Criterion R#21). NOEC values should always be calculated using statistical means. Effect concentrations (ECXX) should be determined by interpolation and not extrapolation. That means that the derived effect concentrations should ideally be within the range of the actually
tested concentrations because extrapolation outside the tested concentration range introduces a great deal of uncertainty. In general, calculated EC10 values that are more than three times lower than the lowest tested concentration are less reliable (Moermond et al., 2016). Test guidelines include recommendations that should be followed for the derivation of effect concentrations, while additional guidance documents provide in depth analysis of statistical considerations in ecotoxicity testing (e.g., OECD, 2006). To handle non-normality and heterogeneity of variance in dose-response ecotoxicity data, alternative regression methods exist not requiring log-transformation and allowing for other family distribution to be used (O´Hara and Kotze, 2010; Ritz and van der Vliet, 2009). For new and existing studies carried out with a suitable experimental design (which allows the calculation of ECXX) EC10, EC20, and EC50 should be reported as a median value together with their 95% confidence intervals. For new and existing studies where the determination of ECXX is not appropriate due to the characteristics of the study design, these effect concentrations should not be reported, and the NOEC should be retained.

The requirement of a dose-response curve (Criterion #22) depends on the objective of the study, for example when the study is performed to test no effects below a certain concentration (limit test) a full dose-response curve is not required. Generally, non-monotonic dose response relationships showing biphasic or bidirectional responses to dose and appearing in U-shaped or inverse U-shaped forms may be problematic. These types of curves are typical for essential metals due to metal deficiency at low exposure concentrations or for some non-essential metals that increase the performance of test endpoints at low doses (hormesis; Bailier and Oris, 2000; Cedergreen et al., 2005). While it is a well-known and accepted phenomenon, hormesis is still problematic for effect concentration derivation (ICCM, 2016). In such
cases, the conventional log-logistic dose response model used to fit the toxicity data requires adaptation (Cedergreen et al., 2005). Alternative regression methods exist (O’Hara and Kotze, 2010; Ritz and van der Vliet, 2009) allowing to explore the existence of hormetic cases (Bailer and Oris, 2000; Cedergreen et al, 2005; Ritz, 2010). For metals, should the resulting L(E)CXX value be below the lowest applied control level (background level) or essentiality level (ICCM, 2016), its reliability/relevance has to be questioned. If no dose-response curve is provided but unbounded values are available, the study can still be valid.

Reliable without restrictions (R1) category should only be assigned when enough raw data is reported so that the effect concentration calculations can be double-checked (Criterion R#23). If statistical methods are poorly described but dose-response curves or tables are reported, the study can still be considered relevant with restrictions as the assessor may derive and/or check effect concentration derivation directly.

CONCLUSIONS AND PERSPECTIVES

This paper proposes a common framework for assessors to evaluate the relevance and reliability of sediment and soil ecotoxicity studies, and provides reporting recommendations for the design and publication of sediment and soil ecotoxicity studies. It builds on the previous CRED method for aquatic ecotoxicity studies and incorporates sediment and soil-specific criteria according to the current practice in risk assessment. This framework is a useful tool for classification of effect data in terms of relevance and reliability and their management during the evaluation process. The most critical decisions for risk assessors in the context of data scarcity are the use of critical data from non-standard test studies, the lack of scientific evidence to consider bioavailability corrections, and the lack of access to original study reports that can
hinder the complete assessment of the available studies. More than aquatic studies, the lack of information in sediments and soil toxicity studies require the use of expert judgement. Where this is the case, judgement should be ideally reviewed with a second assessor. Studies often report data for endpoints and exposure conditions that cannot be categorized to relevant and reliable with/without restrictions categories, which means they cannot be used for regulatory purposes. The reporting recommendations presented here should be used in designing testing strategies for data-scarce substances ensuring that the studies and their results satisfy the evaluation criteria for sediments and soil.

REFERENCES


Table 1. Relevance categories

<table>
<thead>
<tr>
<th>Score</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>Relevant without restrictions: The study is relevant for the purpose for which it is evaluated.</td>
</tr>
<tr>
<td>C2</td>
<td>Relevant with restrictions: The study has limited relevance for the purpose for which it is evaluated.</td>
</tr>
<tr>
<td>C3</td>
<td>Not relevant: The study is not relevant for the purpose for which it is evaluated.</td>
</tr>
<tr>
<td>C4</td>
<td>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.</td>
</tr>
</tbody>
</table>

Source: According to Moermond et al. (2016).
Table 2. Criteria for Reporting and Evaluating Ecotoxicity Data for relevance

<table>
<thead>
<tr>
<th>Biological relevance</th>
<th>1</th>
<th>Is the test species relevant for the compartment under evaluation?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2</td>
<td>Is the test organism relevant for the tested compound?</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>Are appropriate life-stages studied?</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>Are the reported endpoints appropriate for the regulatory purpose?</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>Are the reported endpoints appropriate for the investigated effects or the mode of action of the test substance?</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>Is the exposure duration relevant and appropriate for the studied endpoints and species?</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>Is the effect relevant on a population level?</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>Is the magnitude of effect statistically and biologically significant and relevant for the regulatory purpose (e.g., EC10, EC50)?</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>Are the experimental conditions relevant for the tested species?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Exposure relevance</th>
<th>10</th>
<th>Is the tested exposure scenario, including test substrate, relevant for the species?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>11</td>
<td>Is the tested exposure scenario relevant for the substance?</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>Is the purity of the test substance reported? Or, is the source of the test substance trustworthy?</td>
</tr>
</tbody>
</table>

Source: Adapted from Moermond et al. (2016).
Table 3. Reliability categories

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

Source: According to Moermond et al. (2016).
## Table 4. Criteria for Reporting and Evaluating Ecotoxicity Data reliability

<table>
<thead>
<tr>
<th>Test setup</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Is a standard method (e.g., OECD/ISO) or a modified standard used?</td>
</tr>
<tr>
<td>2</td>
<td>Is the test performed under GLP conditions?</td>
</tr>
<tr>
<td>3</td>
<td>If applicable, are validity criteria fulfilled (e.g., control survival, growth)?</td>
</tr>
<tr>
<td>4</td>
<td>Are appropriate controls performed? Are test item replicates compared to the appropriate controls (e.g., to the solvent control if a solvent was used)?</td>
</tr>
<tr>
<td>5</td>
<td>Were other non-standard effects observed in the controls, like behavioral changes or morphological alterations? Were signs of harm or damage to the organisms in the control checked/reported?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Test compound</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>Is the test substance identified clearly with name or CAS number? Are test results reported for the appropriate compound?</td>
</tr>
<tr>
<td>7</td>
<td>Is the purity of the test substance reported? Or, is the source of the test substance trustworthy?</td>
</tr>
<tr>
<td>8</td>
<td>If a formulation is used or if impurities are present: Do other ingredients in the formulation exert an effect? Is the amount of test substance in the formulation known?</td>
</tr>
<tr>
<td>9</td>
<td>Are there physico-chemical characteristics that may have influenced the behavior of the compound during the study (e.g., solubility, volatility, stability (hydrolysis, photolysis, degradation), log K&lt;sub&gt;OW&lt;/sub&gt;, adsorption) and are they appropriately addressed?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Test organism</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>Are the organisms well described (e.g., scientific name, weight, length, growth, age/life stage, strain/clone, sex if appropriate)?</td>
</tr>
<tr>
<td>11</td>
<td>Are the test organisms from a trustworthy source and acclimatized to test conditions? Have the organisms not been pre-exposed to test compound or other unintended stressors?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Exposure conditions</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>Is the experimental system (e.g., exposure schedule, flow rate/renewal time for sediment toxicity tests, gas exchange for soil toxicity tests, test vessel material, etc.) appropriate for the test substance, taking into account its physico-chemical characteristics?</td>
</tr>
<tr>
<td>13</td>
<td>Is the experimental system (e.g., container size and volume of test sediment/soil, temperature, light/dark conditions (intensity, cycle duration), air humidity for soil toxicity tests, gas exchange/test vessel aeration, feeding) appropriate for the test organism? Have conditions been stable during the test?</td>
</tr>
<tr>
<td>14</td>
<td>Is the test matrix properly characterized and appropriate for the test organism?</td>
</tr>
<tr>
<td>15</td>
<td>Is the method for spiking appropriate (e.g., spiking procedure, equilibration time, equilibration conditions)?</td>
</tr>
<tr>
<td>16</td>
<td>Are a correct range of test concentrations and correct spacing between exposure concentrations applied?</td>
</tr>
<tr>
<td>17</td>
<td>Is the exposure duration defined and appropriate?</td>
</tr>
<tr>
<td>18</td>
<td>Are chemical analyses adequate to verify substance concentrations over the duration of the study?</td>
</tr>
<tr>
<td>19</td>
<td>Is the number or biomass loading of the test organisms in the test system within the appropriate range?</td>
</tr>
</tbody>
</table>

<p>| Statistical design and biological response                                  |                                                                                           |</p>
<table>
<thead>
<tr>
<th></th>
<th>Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>Is a sufficient number of replicates used? Is a sufficient number of organisms per replicate used for all controls and test concentrations?</td>
</tr>
<tr>
<td>21</td>
<td>Are appropriate statistical methods used?</td>
</tr>
<tr>
<td>22</td>
<td>Is a concentration-response curve observed? Is the response statistically significant?</td>
</tr>
<tr>
<td>23</td>
<td>Is sufficient data available to check the calculation of endpoints and (if applicable) validity criteria (e.g., control data, concentration-response curves)?</td>
</tr>
</tbody>
</table>

Source: Adapted from Moermond et al. (2016).