Integrated Environmental Assessment and Management, 2026, **22(3)**, 706–716
https://doi.org/10.1093/inteam/vjaf113
Advance access publication: August 18, 2025
Workshop Synthesis

Special Series

High time to update statistical guidance in ecotoxicology—a workshop synthesis on the revision of OECD document no. 54

Benjamin Daniels[1,*], Thomas Gräff[1], Pia Kotschik[1], and Susanne Walter-Rohde[1]

[1]Section IV, KG I, Ecotoxicology and Bioaccumulation, German Environment Agency, Dessau-Roßlau, Germany

*Corresponding author: Benjamin Daniels. Email: benjamin.daniels@uba.de

Editor's note: This article is part of the special series "Statistical Analysis of Ecotoxicology Data for Regulatory Purposes." The series provides an overview of approaches in statistical ecotoxicology and reflects recent developments, processes, and debates. The papers represent viewpoints from academia, industry, and government.

Abstract

The Organisation for Economic Co-operation and Development (OECD) document No. 54 provides assistance on the statistical analysis of ecotoxicity data to ensure scientifically robust and globally harmonized evaluations of biotests across various regions and regulatory sectors. However, some of the recommended methodologies are outdated due to significant advances in statistical techniques and regulatory requirements. In addition, practitioners have been calling for a more user-friendly structure, aiming to facilitate data analysis for users without extensive statistical expertise. To address these concerns, a research project was initiated by the German Environment Agency with the aim to update OECD No. 54. As part of this project, a dedicated workshop was convened to gather expert perspectives from different sectors (incl. academic, industry, and regulators) on revision needs of OECD No. 54 aiming to better reflect current scientific and regulatory standards. Key debates of the workshop included restructuring the document to improve user accessibility, clarifying terminology, addressing methodological gaps such as assessment approaches for ordinal and count data, and incorporating state-of-the-art modeling approaches for time-dependent toxicity assessment. In addition, the integration of modern statistical practices in hypothesis testing and the provision of clearer guidance on model selection for dose–response analyses were identified as crucial needs for updating OECD No. 54. This synthesis captures the workshop's contributions and recommendations, outlining a roadmap for the revision of OECD No. 54, and highlights the ongoing collaboration with an ISO working group to ensure consistency of standards across regulatory frameworks.

Keywords: OECD No. 54, ecotoxicological data, statistical analysis, workshop synthesis, guideline revision

Background

In environmental risk assessment (ERA), the standardization of statistical analyses for biotest procedures is of particular importance. Establishing international standards in this area ensures robust and comparable results across various studies, thereby enabling accurate assessments of chemical effects on test organisms. Within the European Union, regulations for placing chemicals on the market (e.g., 1107/2009/EC for plant protection products) mandate standardized ecotoxicological laboratory tests. These tests typically follow harmonized test guidelines (TGs) established by the Organisation for Economic Co-operation and Development (OECD). In recent years, statistical aspects of the evaluation of laboratory data have increasingly been outlined separately in each TG.

In 2006, OECD published Document No. 54, titled "Current approaches in the statistical analysis of ecotoxicity data: a guidance to application" (OECD, 2006). This document provides fundamental principles for statistical analyses and describes key statistical methods used to analyze data from laboratory ecotoxicological studies. It not only provides a comprehensive framework but also serves as the basis for the standardized statistical evaluation of OECD TGs.

The current version was developed by eight OECD Member Countries, with France as the lead. The primary objective of this document is to "provide practical guidance on how to analyze observations from ecotoxicity tests." The responsible OECD expert group coordinated its efforts with a working group of the International Standards Organization (ISO) that was simultaneously developing a statistical companion document (ISO, 2006). While OECD TGs serve as a basis for chemical hazard and ERA (prospectively), ISO has traditionally focused on standardization within environmental monitoring and (contaminated) environmental samples (retrospectively). This collaboration aimed to avoid duplication of work while providing harmonized recommendations across both areas.

During the preparation of OECD No. 54 from 2000 to 2006, OECD Member Countries struggled to reach a consensus on how to incorporate specific recommendations regarding statistical approaches. As a result, it was agreed that OECD No. 54 would

Received: May 23, 2025. **Revised:** July 22, 2025. **Accepted:** August 14, 2025

not be designated as an "OECD Guidance Document," but a "Guidance to Application." This distinction reflects the fact that the document does not prescribe obligatory statistical methods for specific purposes or circumstances. Nevertheless, it was recognized as a valuable overview of statistical approaches available at that time.

OECD No. 54 is currently structured around four main chapters, following a brief introduction, a statement of the guidance's scope, and a list of definitions (Figure 1). The first chapter provides an overview of general statistical principles, addressing key experimental design issues such as randomization and replication, along with the data analysis process that includes outlier checks, transformations, and pretreatment considerations. The second main chapter focuses on hypothesis testing, particularly emphasizing no observed effect concentration (NOEC) calculations within the context of multiple tests. The third section details various approaches to dose–response modeling, which are categorized into methods for continuous and categorical data, similar in structure to the hypothesis testing chapter. Finally, the fourth chapter presents biology-based methods derived from mechanistic modeling, explicitly detailing models based on Dynamic Energy Budget (DEB) theory (Nisbet et al., 2000), DEBtox, and DEBtool (Kooijman, 2010).

Some of the approaches described in OECD No. 54 are no longer considered as state-of-the-art. In recent years, new methods have been published and other statistical tests have been adapted to the specific data characteristics in ecotoxicology (see section *New assessment concepts and statistical approaches*). A revision of OECD No. 54 is therefore needed (van Dam et al., 2012; Van Der Hoeven, 2004), as the choice of statistical method and its application has a direct impact on all environment-related OECD TGs and thus on the assessment of the effects of currently regulated chemicals on nontarget organisms.

Workshop objectives and organization

The German Environment Agency (UBA) initiated a project in 2022 aiming to revise OECD No. 54. This project encompasses a systematic literature review and an analysis of feedback from the assessment of revision needs of OECD TGs (Polcher et al., 2023). Additionally, the project includes evaluations of statistical test methods, simulation studies, and power assessments related to various testing approaches. A key component of the project was the establishment of a supporting expert network that would contribute insights through an international workshop (Daniels et al., 2024).

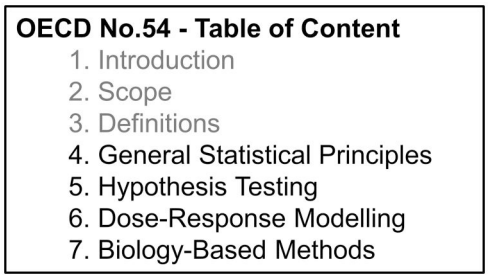

Figure 1. Table of content with main sections of the *OECD No. 54—Current approaches in the statistical analysis of ecotoxicity data: a guidance to application* (OECD, 2006). OECD, Organisation for Economic Co-operation and Development.

The workshop aimed to help identify and discuss the need for revisions of OECD No. 54 from all potential sectors and stakeholders. It took place on September 4 and 5, 2024, at the headquarters of UBA in Dessau. Twenty-four experts from six different European countries participated in person, representing industry, consulting agencies, academia, and regulatory authorities. Additionally, an introductory session featuring presentations on OECD activities and workshop objectives was made available online. Two weeks after the workshop, an online wrap-up event was hold to provide insights and share perspectives regarding the topics discussed during the workshop for online participants all over the world.

Concept of the on-site workshop

To facilitate a lively exchange, the discussion part of the workshop was held exclusively in person in a World Café format, discussing four thematic blocks (see Table 1). Within each block, two discussion rounds for 30 min each at three parallel tables were hold. After the first round, group members rotated tables to encourage a variety of viewpoints and interactions.

At each table, one moderator and one rapporteur were appointed at the beginning of each thematic block to ensure continuity in capturing key insights. At the end of each session, rapporteurs summarized the observed discussions and presented comments from the plenary.

Systematical wrap-up of workshop findings

The rapporteurs' summaries served as a basis for compiling the key issues discussed during the workshop. Experts created 280 notecards during the discussion rounds, which were subsequently organized and categorized on bulletin boards. After the workshop, the information was gathered and compiled for a comprehensive overview of the ideas shared during the sessions. This process allowed the authors to identify and prioritize key topics for a possible update of OECD No. 54.

The following synthesis has been compiled on the basis of the topics discussed during the sessions, aiming to provide an objective overview of the workshop outcome. However, it should be recognized that some issues may still be unintentionally influenced by personal biases. The workshop did not aim to reach consensus, and therefore, not all participants may agree with every aspect of the synthesis presented, including within the team of authors.

Table 1. Thematic blocks of the World Café discussion rounds during the workshop.

Thematic block	Topic	Description
1	Aims and mission of OECD No. 54	Discussion on the overall mission of the document and the main objectives that should be achieved with the revision process.
2	Blind spots of the current document	Which fundamental topics are not mentioned or should be described in more detail in a revised document? How or where can these be addressed?
3	Hypothesis testing	Specific discussion on chapter 5 (hypothesis testing and NOEC [no observed effect concentration] calculation)
4	Dose–response modeling	Specific discussion on chapter 6 (dose–response modeling including EC_X [effect concentration] calculation)

Workshop synthesis and key topics

In the systematic wrap-up, seven main topics were identified, based on the feedback on the notecards, provided by participants (Figure 2). Although the number of notecards per topic does not necessarily correlate with the importance of the topics for participants, Figure 2 provides a general overview of the issues covered.

In the following sections, those seven main topics are summarized, starting with the most frequently raised topic on the notecards: user-friendliness and guidance.

Improving user-friendliness and guidance

Participants agreed that while the current document covers many important statistical and theoretical concepts, it often lacks clear guidance for users trying to navigate its content. There was a strong emphasis on the necessity for clearer instructions on how to analyze data based on specific scenarios, as stated on one notecard: "The guidance is difficult (too difficult) to use in its current form." Users frequently struggle to find relevant advice for evaluating their data when faced with specific data characteristics, such as different scaling types, distributions, or variance patterns. Therefore, it is essential that the recommended methods will be easily accessible to nonstatisticians. To support this, the document should not serve as a statistical textbook but rather support practitioners with practical recommendations tailored to their needs.

A central topic was the call for straightforward navigation. Suggestions included reorganizing chapters and providing step-by-step instructions or decision trees to help users understand complex analyses more easily.

Moreover, protocols for statistical reporting (e.g., by Zuur & Ieno, 2016; Zuur et al., 2010) and updated checklists tailored to different toxicity estimates like NOEC and effect concentration (EC_X) could be included. Additionally, a "quick start guide" explaining key data characteristics, like data scaling and distribution, and subsequent step-by-step procedures for data analysis could be helpful tools. This chapter could be supplemented by a section on visual assessment of data, which is another aspect raised that emphasizes the desire for practical applicability within the guidance.

Flow charts were also proposed, whereby experts stressed that they should be considered as flexible suggestions rather than strict requirements and fixed rules. As a remark on a participant's notecard, "Charts should be recognized as tools for help and not as rules for evaluation." Even if experts had different ideas regarding technical details about enhancing user-friendliness within OECD No. 54, there was clear agreement to improve the document in terms of accessibility and practicability.

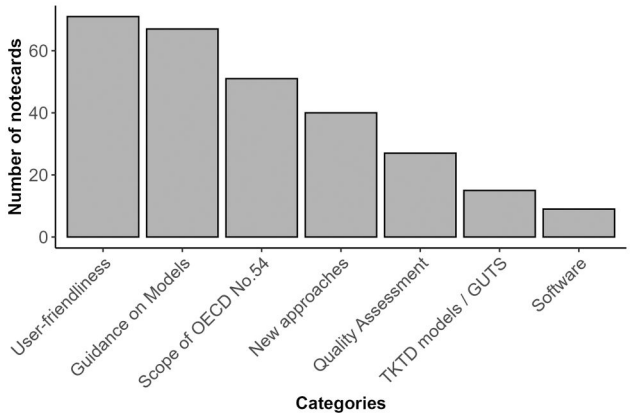

Figure 2. Classification of the prepared notecards at the workshop discussion tables to specific key topics. OECD, Organisation for Economic Co-operation and Development; TKTD models, toxicokinetic-toxicodynamic models; GUTS, General Unified Threshold model of Survival.

Guidance on model selection and test settings

Discussions centered on how to provide effective recommendations for selecting appropriate statistical models and test methods tailored to the type of data being analyzed and the specific characteristics of each study. Participants criticized the current OECD No. 54 for including various methods while making it difficult for users to identify the specific procedures relevant to their data and desired toxicity measures.

Clear instructions are essential to help practitioners make informed decisions. Guidance is needed on when to use dose–response models versus hypothesis testing, as well as the choice of specific testing methods, particularly in complex experimental designs where different statistical approaches may yield different insights. The revision should focus on providing practical advice to enable users to select models that best fit their data and experimental design needs.

Pretreatment and transformation of data

Discussions regarding pretreatment and transformation of data revealed a variety of opinions among participants. While some favored traditional pretests for checking assumptions like normality and homogeneity of variance due to their straightforward nature, this approach was criticized, e.g., for its limitations for small sample sizes, which are common in ecotoxicity studies. Some participants advocated for relying more on expert judgment and "encourage a holistic view" rather than rigid pretest requirements. It was also discussed that TGs could focus on clearly describing how the specific measurement variables in each biotest are scaled and distributed. It was assumed that this should be a valid option for variables measured in many of the longstanding established biotests from TGs.

The relevance of certain recommendations in OECD No. 54, such as routine checks for homogeneity of variances and the application of variance-stabilizing transformations, was debated extensively. Some experts argued that these guidelines are outdated, as modern computational methods provide more efficient and statistically sound alternatives, as shown, e.g., by Bretz et al. (2016) and Hothorn and Kluxen (2020). Others noted that while transformations may still serve practical purposes under specific conditions, they should not be considered a default practice.

Some participants emphasized that transformation should be avoided unless strictly necessary (e.g., to achieve model convergence in certain cases such as negative-binomial or Poisson models) and warned against any pretreatment strategies that could manipulate statistical outcomes. Others emphasized that modern statistical modeling techniques (e.g., Generalized Linear Models [GLMs] and General Linear Mixed Effects Models [GLMEs]) often mitigate the need for transformations altogether (e.g., Szöcs & Schäfer, 2016). It was agreed that transformations should not be performed purely to meet statistical assumptions, as modern statistical approaches offer more robust alternatives.

The consideration of outlier tests was also a significant discussion topic. While this is an important issue for all data types, it was particularly emphasized in the context of endocrine disruptor data. Clear guidance is needed on the selection of outlier tests for evaluating ecotoxicity data, such as Hartley, Dixon-Grubbs, Hampel, and the 3-sigma rule (Hedderich & Sachs, 2020), as well as on justifying the exclusion of outliers. There was unanimous

agreement that outliers should not be removed solely based on test results. They should only be excluded when there are biological reasons to question the plausibility of the results, e.g., technical difficulties with specific replicates may warrant further investigation before exclusion of any data. Overall, establishing clear criteria for handling outliers will enhance the reliability and transparency of statistical analyses in ecotoxicology.

More guidance on hypothesis testing

Participants emphasized the need for straightforward guidance on multiple hypothesis tests for determining NOEC. Fundamental challenges for practitioners were highlighted, such as quickly assessing how data are scaled and which statistical methods are appropriate in different situations (regarding data distributions, number of replicates, sample sizes, etc.). A table with common variables from laboratory tests, such as body weight, growth rates, yield, mortality, immobility, etc., including their scaling, and applicable distribution assumptions, could be included in OECD No. 54 to help practitioners quickly navigate through those issues.

Additionally, there was a strong call for a dedicated section addressing count data, particularly in relation to offspring counts from reproduction tests. Participants noted significant advancements in research and improvements in testing methods over recent years, citing works by Hothorn and Kluxen (2020), Lehmann et al. (2016, 2018), and Szöcs and Schäfer (2015) as valuable contributions to this area. The discussions also revealed a notable gap in current guidance regarding ordinal data analysis. Clear criteria are needed for analyzing ordinal outcomes commonly encountered in biological assessments, such as histopathology or visual evaluations of plant injury. Recently developed methods like the Rao-Scott Cochran-Armitage by Slices trend test (RSCABS; Green et al., 2014) and the multiquantal Jonckheere-Terpstra (MQJT, Green et al., 2018) were mentioned as potential tools that could be integrated into the guidance. However, it was acknowledged that evaluating ordinal data in the context of visual plant injury has not yet been fully resolved within the scientific community; thus, integrating an approach for OECD No. 54 may be premature given that several questions remain open regarding issues of data quality, including subjectivity, variability, and reproducibility.

A debate was also held on recommendations for one-sided versus two-sided statistical testing for ecotoxicity data. Some participants argued that the direction of an adverse effect should be established prior to conducting any statistical analysis. It was suggested that clarity regarding test direction should be provided across all TGs since typical test systems are generally designed in terms of detecting a certain effect direction in the treatment compared to the control. In most cases, a decrease of the measured variable compared to the control is targeted; however, increases may also be relevant biological effects and their evaluation is required in several TGs (e.g., OECD 216, 2000). OECD No. 54 should include specific examples from typical biotests and their measured variables to clarify when each test direction should be preferred.

Guidance on dose–response modeling

A strong need for updated guidance on nonlinear regression models as part of the revision of dose–response modeling approaches was identified. OECD No. 54 should incorporate recent methodological advancements, particularly emphasizing the features of the drc package (Ritz et al., 2015) from R software (R Core Team, 2024).

A key point of discussion was the potential replacement of Slob's "nested nonlinear models" with more commonly used alternatives such as lognormal, log-logistic, and Weibull models featured in the drc package. While some participants argued that Slob's models still hold relevance, others questioned their value, noting that they essentially represent Weibull Type 1 functions with slightly different parameterization (Ritz, 2010). In terms of clarity and usability, it was suggested that focusing on widely accepted models would be more beneficial.

These discussions underscored the necessity for clearer guidance on selecting appropriate models. A critical aspect involved determining when to use two-, three-, or four-parameter models. Participants emphasized that the choice of model should be guided by biological plausibility rather than mere statistical convenience. This is particularly relevant for using four-parameter regressions where the lower limit is not fixed at zero; applying these models to continuous or count data from ecotoxicological tests is only justified if deviations from a nonzero lower limit can be explained based on biological or substance-related reasons (Daniels et al., 2026).

Another gap is the distinction between EC_X (effect concentration) and ICx (inhibition concentration). Currently, OECD No. 54 does not differentiate, unlike the statistical guidance document from Canada (Government of Canada, 2007) which explicitly addresses this issue. This lack of clarity leads to confusion regarding the definition of EC_X for continuous variables in OECD No. 54, particularly in cases where dose–response curves have nonzero bottom parameters, such as those generated by four-parameter functions in the drc package or a three-parametric model with offset that do not fix the lower limit at zero. Some participants argued that regulatory EC_X values should be defined as a percentage of inhibition compared to control conditions, suggesting that lower limits could be disregarded when calculating EC_X values (i.e., absolute EC_X). However, this was criticized by others as this approach aligns more closely with the definition of ICx rather than EC_X (Government of Canada, 2007) and is also inconsistent with how Slob outlined the derivation of EC_X values in his work (OECD, 2006). Clarifying this issue is essential within the context of revising OECD No. 54 (Daniels et al., 2026), especially since benchmark dose (BMD) calculations are increasingly utilized as alternatives to traditional EC_X calculations (EFSA Scientific Committee, 2022). The BMD represents a dose level associated with a specified change in response relative to a control group and corresponds more closely to ICx (or "absolute EC_X") than to traditional EC_X values. Participants noted that clearer guidance is necessary to address these discrepancies, particularly concerning model selection for regulatory purposes.

In addition to these fundamental questions regarding the revision of dose–response modeling and EC_X calculations within OECD No. 54, several detailed issues surrounding regression methodologies were identified, e.g., how to best derive EC_X values from one or more curve regressions. Guidance is needed for the selection of one out of multiple dose–response models, especially if all of them perform equally well. There was also discussion about possible benefits of mandate testing and reporting a fixed set of models. The concept of model averaging from all acceptable fits was introduced as a possible approach. Other options included choosing the most conservative EC_X, the model with the lowest confidence intervals, or the 90% quantile. However, most participants agreed that it should not be mandatory to test and report a specific number of models in all cases, particularly when dealing with straightforward data situations where such requirements may not justify the effort involved. Instead, it was

recommended to generally test various functions and select the most trustworthy EC_X based on different curve parameters, including goodness-of-fit testing, the Akaike Information Criterion, and reliability criteria (EFSA, 2019).

Participants further deliberated the need of established quality requirements for datasets used in dose–response modeling. This discussion paralleled conversations about hypothesis testing and also raised very specific questions of, e.g., incorporating Poisson models for count data analysis. Another important area was the handling of steep dose–response curves with missing data in the critical area of interest. Clearer guidance is needed within OECD No. 54, including commonly used practices such as calculating geometric means for continuous data or interpolation methods (e.g., Spearman Kaerber or Moving averages) for quantal data. Additionally, participants expressed a desire for an explanation on how to manage control (0) values when working with log-transformed doses. It was suggested to recommend the use of very low replacement values, ensuring these values fall below typical treatment levels while remaining within the plateau region of the curve to avoid an influence on the EC_X calculation.

Lastly, discussions highlighted inconsistencies in how mortality is treated in dose–response models assessing sublethal effects. It was pointed out that mortality is a distinct dichotomous variable (alive/dead) that should not be conflated with continuous variables used for analyzing sublethal endpoints, such as by setting those affected replicates equal to zero. Given that dead individuals can still impact sublethal endpoint results (especially if replicates are dropped despite an observed effect), careful planning is essential to minimize excessive mortality and reduce data distortion. Some experts recommended explicitly accounting for mortality data within multifactorial statistical models to prevent misinterpretation of sublethal endpoints; however, this suggestion faced criticism due to concerns over potentially exceeding practitioners' statistical capabilities.

Scope of OECD No. 54

There is a need to critically evaluate the scope of OECD No. 54 in light of evolving scientific and regulatory demands. Participants debated whether to maintain the existing scope or expand it to incorporate new priorities that have emerged in recent years. Currently, the document focuses on statistical methods for analyzing standardized univariate ecotoxicity test data from laboratory settings, including the derivation of toxicity estimates like EC_X and NOEC.

A significant topic was the role of protection goals in shaping statistical evaluations. While some participants argued that statistical evaluation should remain independent from protective risk assessment objectives—stating that "the best statistics for biological data should be recommended"—others contended that understanding the intended protection goals is essential for making informed decisions about statistical methods. For instance, while NOEC may not be statistically optimal for interpreting toxicological data, it remains a regulatory requirement, necessitating guidance on how to calculate it properly.

There was considerable interest in extending the guidance to cover more complex testing scenarios, such as (semi-) field tests and species sensitivity distributions, in addition to its focus on laboratory data. However, several experts expressed concerns about user-friendliness regarding this expansion of scope. They cautioned that incorporating too many topics could result in a document that is too complex and lacks clarity for practitioners focused on laboratory test evaluations. As one expert noted, there is a strong desire to generate precise and straightforward guidance for laboratory tests specifically, without overwhelming users with excessive information. Nevertheless, the statistical evaluation of higher-tier tests, such as field tests, is an urgent issue that requires further guidance.

Another significant issue is the potential need for a dedicated detailed chapter on experimental design (DoE), as improved guidance on statistically designing biotests to ensure reliable outcomes is lacking. This includes providing specific advice on how various experimental designs influence statistical power and overall data quality. One participant pointed out, "The choice between assessing a NOEC and estimating an EC_X should actually be made before designing the study." In this context, the number of replicates in an EC_X approach is less critical than the amount and selection of treatments; conversely, replicate design plays a crucial role in determining the power of NOEC assessments. It was agreed that considering statistical aspects prior to developing a biotest is at least as important as conducting post hoc evaluations. However, some participants cautioned that maintaining a clear focus on supporting users with laboratory test evaluations is essential, suggesting that delving into biotest design could detract from this primary objective.

Some notecards reflected these discussions, e.g., by suggesting to add multivariate approaches and establishing links between experimental design and statistical analyses while cautioning against compromising user-friendliness through excessive complexity. In synthesis, although opinions varied regarding how best to define the scope of OECD No. 54, reflecting diverse perspectives among participants, there was general agreement on updating this aspect to ensure its relevance and applicability in current ecotoxicological assessments. Balancing user-friendliness with comprehensive coverage will be crucial as revisions progress.

New assessment concepts and statistical approaches

Numerous ideas were proposed regarding new assessment concepts and statistical approaches. The inclusion of these ideas in OECD No. 54 will depend on various factors, such as the need to keep the document straightforward and user-friendly, the primary focus on established methodologies in ERA, and ensuring that recommended procedures are accessible to users with varying levels of statistical expertise. Additionally, international applicability of methods will be a key consideration, alongside the available resources in the revision process and finding common ground among member countries.

Regardless of the specific proposals for new approaches that will be outlined in the following sections, participants consistently emphasized that any new methods must undergo thorough validation and testing using both simulated and real datasets before being widely adopted for regulatory statistics. The guidance should stress the importance of selecting appropriate statistical methods that are fit for purpose and align with the specific objectives of ecotoxicological studies. This involves a clear understanding of endpoint characteristics, such as ecological relevance, and recognizing the underlying assumptions associated with different statistical techniques. By doing so, OECD No. 54 can better support practitioners in making informed decisions based on robust analytical frameworks.

The family of generalized linear models

Some feedback focused on the need to incorporate more recently applied statistical approaches, particularly GLM and related concepts. These frameworks offer flexible handling of various data types and structures, including random effects and nonnormal distributions, making them highly relevant for ecotoxicological studies. They can effectively address variable transformations

and other technical challenges. Importantly, these methods have been accessible to the broader research community for several decades and are more and more used for the evaluation of ecotoxicity studies (e.g., field tests) (Vermeiren et al., forthcoming).

Participants recommended including specific information and practical applications regarding various model families such as GLMs, General Additive Models (GAMs), Linear Mixed Effects Models (LMEs), and GLMEs. These models were frequently cited during discussions as essential tools in ecotoxicological assessments. However, it was noted that while GLMs are generally very useful, they are often misapplied. This issue may not be evident from the typical result tables presented; rather, it becomes apparent only through visual verification of GLM predictions against actual data.

Experts suggested that researchers should not default to use standard GLMs but instead carefully select the appropriate model by assessing residual distributions. In summary, participants emphasized the importance of always checking residual distributions, often found to be nonnormally distributed, and visually verifying GLM predictions alongside measured data.

However, the characteristics, strengths, and application recommendations for these models also reveal significant weaknesses concerning standardized use in TGs. As discussed in the workshop, recommendations for tests from the GLM(E) family were met with considerable skepticism regarding their applicability in standardizing biotest procedures and regulatory frameworks. It was pointed out that the wide range of options available within GLMs complicates comparability across studies. One notecard succinctly stated: "Complexity of GLMs—so many methods to get *p*-values. How to standardize?"

One participant illustrated this balance between pros and cons effectively: "PROS: high level of statistical analysis; make best use of data; can be fitted best to study design; CONS: difficult to analyze; high complexity; hard under Good Laboratory Practice." Thus, integrating this family of methods into TG standards remains uncertain even after the workshop discussions. A potential solution could involve proposing their use specifically related to biotests under clear guidance and defined limits, such as specifying assumptions about data distribution and identifying which influencing factors (e.g., specific environmental parameters or plot designs) should be considered in modeling decisions (random or fixed effects, nested variables, etc.).

Equivalence testing—proof of absence of effects

The use of equivalence tests for the standardized evaluation of biotests was a notable topic. Advocates for this approach argued that, in ecotoxicology, a fundamentally different methodology than the currently standard hypothesis testing would be more appropriate. Based on the precautionary principle, the use of chemicals should only be accepted if there is evidence that there are no adverse effects present. Therefore, using a significance level of $\alpha = 0.05$, which is roughly interpreted as accepting that effects exist only with a 95% confidence, may not be suitable, e.g., under current practices, a calculated *p*-value of .10 might lead to the conclusion that there is no statistically significant difference between the control and a treatment. However, this could mask the fact that the probability of an actual effect is quite high in such a case. Instead, the focus should be on avoiding acceptance of substances for which it has not been demonstrated (with high confidence) that they do not cause any effects. This shift in perspective prioritizes minimizing Type II errors (focusing on β rather than α). Cohen (2013) proposed that the maximum acceptable probability of a Type II error should be 20%. One suggested approach involves formulating an inverse null hypothesis: rather than hypothesizing that there is no effect unless proven otherwise, we could assume effects are present unless demonstrated otherwise.

Some experts expressed reservations for the application of equivalence testing to laboratory test evaluations. They argued that reversing hypotheses may be acceptable in higher-tier field tests, where substances are already suspected of posing unacceptable risks. This method could lead to excessive Type II errors in standardized tests due to high variability in some datasets (Hotopp et al., 2024). Such errors would imply accepting the null hypothesis of risk when it may not accurately reflect reality, thereby complicating the realistic interpretation of ERA results. One participant succinctly captured this concern with a note card comment: "Equivalence testing should be 'road tested' before being recommended," indicating a need for further research on this topic. However, the concept of equivalence testing has gained increased attention in ecotoxicology recently due to its incorporation into guidance documents (EFSA, 2023a).

Alternative toxicity metrics

Discussions regarding alternative toxicity measures to the stated NOEC and EC_X metrics received limited attention, likely due to the primary focus on updating existing concepts. However, participants acknowledged that alternative metrics have been proposed and should be considered during the revision process. The challenge lies in balancing which new approaches or expansions of existing concepts genuinely add value versus those that could complicate practical application by introducing yet another method without clear benefit.

Examples of alternative metrics discussed included BMD for birds and mammals (EFSA Scientific Committee, 2022), which may require different experimental designs compared to those outlined in OECD TGs. It is important to clarify that BMD can be equated with EC_X when EC_X is defined as a percentage of inhibition relative to control conditions (i.e., absolute EC_X). Additionally, calculations for NEC (no effect concentration) and NSEC (no significant effect concentration) derived from nonlinear models were mentioned as conceptionally comparable (Fisher & Fox, 2023; Fisher et al., 2024; Pires et al., 2002), but distinct from NEC calculations obtained through mechanistic models as currently described in Chapter 7 of OECD No. 54. This distinction is crucial to avoid confusion among practitioners regarding the applicability and interpretation of these different methodologies. While alternative toxicity metrics were not the central focus of the workshop discussions, their potential inclusion in the revised guidance warrants careful consideration. This is to ensure that they enhance, rather than detract from, the usability and clarity of OECD No. 54.

Bayesian approaches

While Bayesian approaches were only mentioned briefly during discussions, indicating their practical implementation remains limited to date, they were recognized as valuable tools worth exploring further within OECD No. 54 revisions. While Bayesian methods are often favored for their ability to incorporate prior information and explicitly address uncertainty, prior information is not always available in practice.

Besides a brief description of frequentist and Bayesian methods, clear guidance on their effectiveness would also be useful. Moreover, documentation of any prior information from, e.g., historical control data (HCD) or expert judgment is essential for standardization of those methods. In addition, all assumptions made during modeling should be recorded for transparency and reproducibility. The guidelines should outline estimation procedures (e.g., the MCMC methods Metropolis-Hastings, Gibbs

sampling, and STAN) as well as methods for assessing convergence, including burn-in time and convergence metrics.

Participants recommended that there should be clear and distinct explanations that nearly all methods can also be implemented under a Bayesian framework. Nevertheless, explanations of when Bayesian methods may have limitations compared to frequentist approaches should be included in the revision of OECD No. 54. For many experts, it would be beneficial to state that both methodologies are valid without showing preference for one over the other. However, as these concepts require some prior knowledge and are not immediately applicable to inexperienced laypersons, it is still open for discussion among OECD member countries whether Bayesian approaches can be considered ready for use in this context.

Quality assessment and diagnostics of tests and methods

Throughout the workshop, the importance of assessing the methodological quality of statistical approaches was a recurring issue across all topical discussion rounds (Table 1). Participants emphasized the need for clear guidelines on how to evaluate whether a model or test is effective and what conclusions can be drawn from the results. A strong call was made for guidance on comparing different analytical methods, such as various multiple hypothesis tests for NOEC calculations or regression models and functions for EC_X determinations. Many participants expressed a desire for straightforward recommendations, ideally presented in decision trees or similar formats to aid decision-making.

It is crucial to differentiate between quality criteria applicable to hypothesis testing related to NOEC derivation and those relevant for regression analyses in dose–response modeling (i.e., EC_X calculation). Regardless of the method employed, be it hypothesis testing, regression analysis, or TKTD modeling, the ability to compare models and assess their effectiveness remains vital.

Participants frequently highlighted that uncertainty and variability must be considered alongside test power and modeling results. In contexts involving hypothesis testing, metrics like the minimum detectable difference (Brock et al., 2015; Duquesne et al., 2020; Hotopp & Fuelling, 2019), minimum significant differences, or confidence intervals of responses (Mair et al., 2020) were identified as tools for post hoc power analyses. However, it was also suggested in the workshop that the guideline should not simply list all the methods, but should identify one method which would then be consistently reported throughout for comparability. One expert noted that consistent minimum levels of statistical power should be maintained and reported across all testing methods, stating that from an ecological point of view, "the requirements for statistical power apply regardless of endpoint variability." This perspective highlights the importance of robust standards to ensure that practitioners can evaluate their chosen methods effectively, independently of the tested endpoint.

For dose–response modeling, criteria such as visual residual checks of models, goodness-of-fit tests, and Akaike Information Criterion (AIC) for comparing models were recommended. Specific methods such as interpreting and using confidence bands, employing lower 95% limits as toxicity estimates, and utilizing existing reliability criteria for direct comparisons of normalized widths and reliability of EC_X values (European Safety Authority, EFSA et al., 2019) were also deemed necessary. These practices help to ensure robust and meaningful assessments.

TKTD models and GUTS

Due to developments over the past two decades, particularly in time-dependent toxicity (TKTD [toxicokinetic-toxicodynamic]) modeling as supported by the GUTS (General Unified Threshold model of Survival) framework (Jager et al., 2011), the existing chapter on biology-based methods is outdated and does not align with modern requirements, especially for ERA (e.g., Baas et al., 2018).

Some experts suggested completely removing this chapter, arguing that comprehensive textbooks (Jager & Ashauer, 2025; Kooijman, 2000) and guidelines on good modeling practices from authorities like the EFSA (EFSA, 2014) already describe the use of TKTD models sufficiently. One participant noted, "A further general description of the use of TKTD models is not necessary." Additionally, it was emphasized that OECD No. 54 serves as a statistical guidance document, and including mechanistic approaches could distract from its core scope. It also has been shown that mentioning specific models such as the DEB model in the OECD No. 54 has not led to significant adoption of these methods in ERA (Becker et al., 2024).

However, there were also voices advocating for retaining individual-level TKTD models while integrating them into OECD No. 54. A key proposal was to present these models directly as alternatives to traditional regression methods within the dose–response modeling chapter. This integration would provide users practical recommendations for applying TKTD models and show their potential alongside conventional statistical approaches.

One significant advantage of incorporating TKTD models into dose–response analyses is their ability to derive time-independent EC_X values (calculate an $EC_X(t)$ for any x, at each time point t), if sufficient data are available (Charles et al., 2022). This capability allows for a more nuanced understanding of how substances affect biological systems over time, which can be particularly valuable in ERA when temporal dynamics play a crucial role. Supportive notecards reflected this sentiment by suggesting including readily usable TKTD models like GUTS for test evaluations and proposing that "TKTD should be integrated into other chapters (e.g., DR modeling)." There were calls to refine the document's structure by merging biological-based methods with statistical methodologies used to derive regulatory endpoints.

Another point raised by experts highlighted the diversity of mechanistic models available, from individual-level toxicity to spatially explicit population modeling, and cautioned against discussing each model in detail due to their variety. Instead, they recommended establishing general criteria applicable across all mechanistic approaches like "include variability and uncertainty" and "choose lower confidence limits when determining endpoints." This approach would allow users to understand how results can take account of uncertainty without being lost in non-representative examples.

While opinions varied regarding the relevance of biology-based methods within OECD No. 54, there were a number of comments on enhancing their integration rather than excluding or superficially addressing these important methodologies in ecotoxicological assessments. By positioning TKTD models as alternatives within dose–response modeling discussions, users could benefit from a broader toolkit for analyzing ecological data effectively.

Software use

The workshop discussions highlighted the need for a comprehensive review of the software recommendations in OECD No. 54 as many of already included recommendations are outdated and linked to commercial software, which should be removed to align with contemporary OECD standards. It was agreed that outdated software references should be removed and replaced with a focus on available methods, including their pros and cons.

A lively debate emerged regarding whether specific software should even be mentioned. Noncommercial software often lacks proper validation, and R packages can evolve rapidly, making any specific recommendations quickly obsolete. One participant suggested to link the software section to a "living document" website as an extension of the guidance to keep noncommercial recommendations up to date instead of listing specific programs directly. However, concerns about how this would fit into the OECD framework remained unresolved.

Despite these challenges, there was general agreement on establishing quality standards for software use if no particular programs are named. Experts emphasized that unvalidated software should not be used for standardized analyses and that many academic or custom-built tools (like those developed in R) often lack proper validation. Tools that allow users to individually modify fundamental procedures without oversight will pose significant risks, especially for users without extensive statistical knowledge who may inadvertently introduce errors into their analyses. It was also considered essential to maintain version control when reporting results. One expert summarized these issues stating that "It is crucial to verify software used in analyses, as well as to document and report methodologies to ensure reproducibility and reliability."

If, on the other hand, specific software is mentioned in the document, experts require that "guidance should include R, RStudio, and various relevant packages like *morse* or *drc*," while also noting useful online tools such as PROAST (RIVM, 2012) or BMDS (USEPA 2000). Although there were differing views on the most appropriate way to handle software recommendations in OECD No. 54, there was a broad consensus that this part of the guidance document needed to be updated.

Additional aspects and general remarks of revision

Several important topics were also discussed during the workshop. While a detailed analysis exceeds the scope of this summary, some key themes deserve mention:

1. One critical point raised was that NOEC as a toxicity metric must continue to be addressed within OECD No. 54. Although there has been longstanding criticism regarding its use, dating back to a 1998 OECD recommendation suggesting it should be phased out (OECD, 1998), it remains prevalent in regulatory applications today. Many practitioners expressed frustration at being forced to use NOEC despite its criticisms; thus, guidance on NOEC approaches remains essential in this revision process. While it may not be feasible to eliminate NOEC from this guideline entirely, addressing critiques surrounding its use and presenting arguments for both sides will help clarify its role moving forward. Ultimately, phasing out NOEC must occur through regulatory legislation.
2. Another noteworthy topic was the necessity of Good Laboratory Practice (GLP) standards for statistical methods. Participants discussed the implications that such standards could have for regulatory compliance and data integrity. It was generally agreed that GLP standards pertain more to the GLP framework itself rather than being part of the statistical guidance in OECD No. 54. The discussion highlighted the need to consider whether statistical appropriateness should be included alongside documentation requirements as part of GLP standards.
3. Another area of debate involved historical control data. There was no consensus on how HCD should be integrated into future investigations, with one participant questioning, "Historical control data to decide on distribution???". The relevance of HCD in the context of OECD No. 54 was emphasized, especially given new recommendations and publications from EFSA (2023b) and Kluxen et al. (Kluxen et al. 2021), which could inform updates to OECD No. 54. If the use of HCD is integrated, the data must be carefully checked to ensure that current data and HCD were measured under the same conditions and originate from the same population. However, concerns have been raised as to whether HCD can ever truly originate from the same population, neither in a biological nor in a statistical sense. Coja et al. (2022) define HCD as data "observed in the same species and the same type of toxicity study as the study under evaluation," which helps to define minimum criteria for inclusion, but does not fully resolve concerns about biological and statistical comparability. In practice, HCD are often part of meta-analyses rather than a standardized approach for analyzing biotests outlined in TGs, as data are normally gathered from different test and test setups.
4. General remarks were also made regarding the broader context of statistical guidance within an international framework. While there is a desire to focus on robust statistical methodologies, participants acknowledged that different needs from Member States must be considered. One key requirement is ensuring that statistical test methods are accessible even to nonstatisticians using straightforward tools. This requires prioritizing general applicability over cutting-edge techniques while still providing meaningful results.

The dilemma between standardization and scientists' drive for innovation

In revising OECD No. 54, we will face a significant dilemma that highlights the tension between the need for standardization and the desire of many statistical experts for innovation, which also became apparent during the workshop. Effective statistical analysis often requires adapting to the specific features of a dataset, which sometimes requires case-by-case decisions. However, this flexibility can clash with the goal of creating guidance that imposes strict and straightforward requirements for statistical evaluation. It is inherently challenging to capture the wide variety of data situations within a single framework in a clear and meaningful way.

In this context, it is important to consider the targeted audience of the document: the guidance should be explicitly helpful and applicable for nonstatisticians with an ecotoxicological background and only basic statistical knowledge. A key goal of the revision is to ensure that recommended methods are accessible to all stakeholders while clearly defining areas where new methods can be applied. As an international guidance document, OECD No. 54 must be globally applicable, striking a balance between practicality and complexity that is appropriate for different contexts. The objective stated in 2006 to "provide practical guidance on how to analyze observations from ecotoxicity tests" emphasizes practicality over striving for "state-of-the-art" methodologies.

Finding a balance between providing straightforward guidance where it is helpful and allowing room for innovative ideas is crucial. This balancing act becomes particularly complex in ERA, where standardization plays an important role in ensuring consistency across assessments while still allowing for the variability inherent in ecological data.

Outlook on the revision process following the workshop

The key topics identified above provide a first roadmap for the next phases of a revision process of the OECD No. 54. How can we ensure that the identified needs and options for revision could be effectively integrated into a future version of the document? Figure 3 illustrates a possible procedure and provides a link to the formal process of draft development at the OECD level.

The first step in the revision must be a final decision on the scope of OECD No. 54 and the subsequent key objectives of the revision, in particular, whether the focus should remain on recommendations strictly limited to the evaluation of laboratory tests or whether the scope should be broadened to include semi-field and field tests. Definitions and terminology will be updated to ensure clarity and consistency throughout the document.

The following structural revision aims for more practical and user-friendly guidance, beginning with an updated introduction designed as a quick guide for users navigating from data visualization to toxicity estimation based on their specific data types and questions. This sets the stage for drafting focused on reporting, communication (including recommendations for quality assessment and diagnostics of test procedures; see above), statistical methods, and integrating supporting tools. Given the wealth of ideas shared during the workshop, it may not be feasible to incorporate all amendments into the first draft document within available time constraints. Tasks requiring collaboration among experts depend significantly on their workload capacities.

Germany, being the lead of the OECD project, will prepare an initial draft document. This draft will be reviewed by an OECD expert group designated for that task and consisting of nominated experts from OECD member countries, industry, and non-governmental organizations (NGOs). This review process is coordinated by the Working Party of National Coordinators of the OECD Test Guidelines Programme (WNT) and the OECD Secretariat. Once a revised draft is available from the Expert Group, a WNT commenting process takes place in which the National Coordinators, industry, and NGOs submit comments and feedback. The focus here is on achieving the best possible international standardization and at the same time integrating as many aspects as possible that were developed in the workshops.

A crucial organizational aspect is the collaboration with ISO to achieve a consistent revision of the sister document, the ISO/TS 20281 (ISO, 2006). It is essential for stakeholders and secretariats of both OECD and ISO to maintain continuous communication throughout the updating process, aiming to merge revision processes and review formats wherever possible. This cooperation within the ISO context requires careful planning and monitoring, as OECD and ISO operate under different governance principles, business models, and objectives. The collaboration offers a valuable opportunity to ensure a coherent approach to ecotoxicological statistical evaluation on a global scale.

The workshop provided valuable insights, reinforcing that an update of OECD No. 54 is indeed timely and represents a crucial step toward harmonizing the statistical evaluation of ecotoxicity data using contemporary methods and approaches. Specific areas for revision were identified, which should be addressed through the collaborative efforts of member states and involved stakeholders.

However, many of the discussed aspects still require further exploration within the scientific community, as some points have only been touched upon without reaching a consensus at this stage of the discourse. The ongoing process will continue to engage our community in the coming years, and we look forward to these discussions shaping the future of ecotoxicological assessments.

Data availability

Workshop participant statistics available on request to authors.

Author contributions

Benjamin Daniels (Conceptualization, Formal analysis, Methodology, Project administration, Visualization, Writing—original draft, Writing—review & editing), Thomas Graeff (Writing—original draft, Writing—review & editing), Pia Kotschik (Conceptualization, Project administration, Writing—original draft, Writing—review & editing), and Susanne Walter-Rohde (Conceptualization, Writing—original draft, Writing—review & editing)

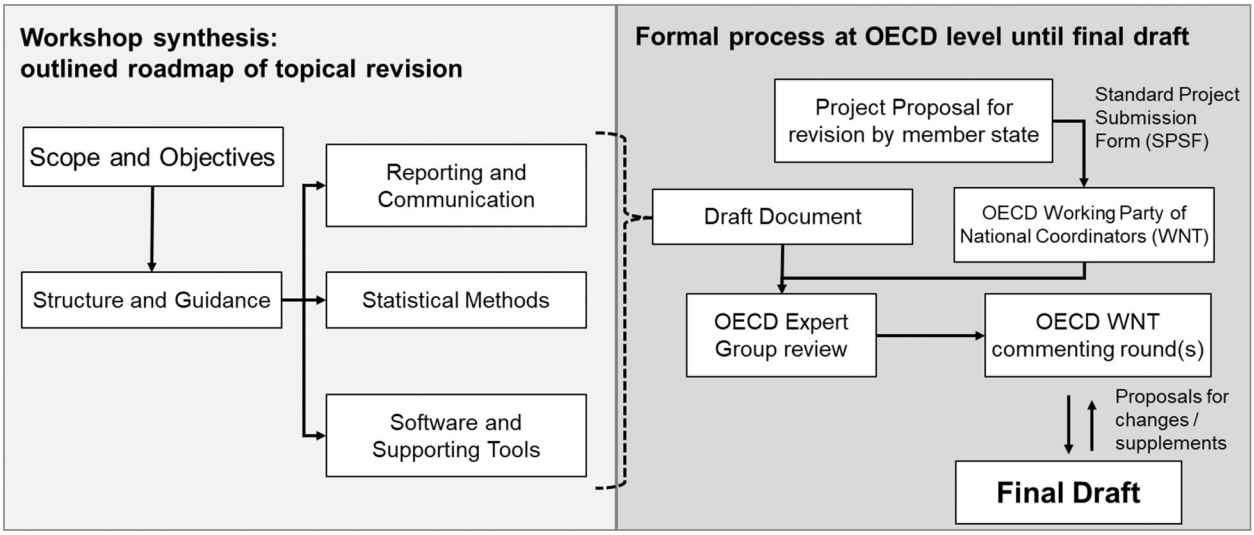

Figure 3. Roadmap for revising the OECD No. 54 and formal process at the OECD level. OECD, Organisation for Economic Co-operation and Development.

Funding

Ths work was funded by the German Environment Agency.

Conflicts of interest

B.D. works on a part-time basis for ToxRat Solutions, which offers commercial software for the statistical evaluation of standardized ecotoxicological tests. The authors declare no conflicts of interest.

References

Baas, J., Augustine, S., Marques, G. M., & Dorne, J.-L. (2018). Dynamic energy budget models in ecological risk assessment: From principles to applications. *Science of the Total Environment*, *628–629*, 249–260. https://doi.org/10.1016/j.scitotenv.2018.02.058

Becker, J., Liess, M., Kramer-Schadt, S., Franz, M., & Jager, T. (2024). Critical evaluation of effect models for the risk assessment of plant protection products. *Umweltbundesamt Texte*, *41*, 596. https://www.umweltbundesamt.de/publikationen/critical-evaluation-of-effect-models-for-the-risk

Bretz, F., Hothorn, T., & Westfall, P. (2016). *Multiple comparisons using R*. Chapman and Hall/CRC. https://doi.org/10.1201/9781420010909

Brock, T. C. M., Hammers-Wirtz, M., Hommen, U., Preuss, T. G., Ratte, H.-T., Roessink, I., Strauss, T., & Van den Brink, P. J. (2015). The minimum detectable difference (MDD) and the interpretation of treatment-related effects of pesticides in experimental ecosystems. *Environmental Science and Pollution Research International*, *22*, 1160–1174. https://doi.org/10.1007/s11356-014-3398-2

Charles, S., Ratier, A., Baudrot, V., Multari, G., Siberchicot, A., Wu, D., & Lopes, C. (2022). Taking full advantage of modelling to better assess environmental risk due to xenobiotics—The all-in-one facility MOSAIC. *Environmental Science and Pollution Research International*, *29*, 29244–29257. https://doi.org/10.1007/s11356-021-15042-7

Cohen, J. (2013). *Statistical power analysis for the behavioral sciences* (2nd ed.). Routledge. https://doi.org/10.4324/9780203771587

Coja, T., Charistou, A., Kyriakopoulou, K., Machera, K., Mayerhofer, U., Nikolopoulou, D., Spilioti, E., Spyropoulou, A., Steinwider, J., & Tripolt, T. (2022). Preparatory work on how to report, use and interpret historical control data in (eco)toxicity studies. *EFSA Supporting Publications*, *19*, EN-7558. https://doi.org/10.2903/sp.efsa.2022.EN-7558

Daniels, B., Graeff, T., Walter-Rohde., & Kotschik, S. (2024). *Activities to revise the OECD Document No. 54 (and ISO/TS 20281) on statistical analysis of ecotoxicity data [Conference presentation]*. SETAC Europe 34th Annual Meeting, Seville, Spain.

Daniels, B., Hommen, U., & Ratte, M. (2026). Four-parameter nonlinear regression and maximum achievable effect in ecotoxicology: Just visually appealing or relevant for risk assessment? *Integrated Environmental Assessment and Management*, *22*, 734–740. https://doi.org/10.1093/inteam/vjaf085

Duquesne, S., Alalouni, U., Gräff, T., Frische, T., Pieper, S., Egerer, S., Gergs, R., & Wogram, J. (2020). Better define beta-optimizing MDD (minimum detectable difference) when interpreting treatment-related effects of pesticides in semi-field and field studies. *Environmental Science and Pollution Research International*, *27*, 8814–8821. https://doi.org/10.1007/s11356-020-07761-0

EFSA Scientific Committee. (2022). Guidance on the use of the benchmark dose approach in risk assessment. *EFSA Journal*, *20*, e07584. https://doi.org/10.2903/j.efsa.2022.7584

European Food Safety Authority (EFSA). (2014). Scientific Opinion on good modelling practice in the context of mechanistic effect models for risk assessment of plant protection products. *EFSA Journal*,*12*(3), 3589, 92 pp. https://www.efsa.europa.eu/en/efsajournal/pub/3589

European Food Safety Authority (EFSA). (2019). Outcome of the Pesticides Peer Review Meeting on general recurring issues in ecotoxicology. *EFSA Supporting Publications*, *16*, 1673E. https://doi.org/10.2903/sp.efsa.2019.EN-1673

European Food Safety Authority (EFSA). (2023a). Revised guidance on the risk assessment of plant protection products on bees (*Apis mellifera*, *Bombus* spp. and solitary bees). *EFSA Journal*, *21*, 7989, 133. https://www.efsa.europa.eu/en/efsajournal/pub/7989

European Food Safety Authority (EFSA). (2023b). Guidance on the risk assessment for birds and mammals. *EFSA Journal*, *21*, 7790, 300. https://doi.org/10.2903/j.efsa.2023.7790

Fisher, R., & Fox, D. R. (2023). Introducing the no-significant-effect concentration. *Environmental Toxicology and Chemistry*, *42*, 2019–2028. https://doi.org/10.1002/etc.5610

Fisher, R., Fox, D., R., Negri, A., P., van, Dam, J., Flores, F., & Koppel, D. (2024). Methods for estimating no-effect toxicity concentrations in ecotoxicology. *Integrated Environmental Assessment and Management*, *20*, 279–293. https://doi.org/10.1002/ieam.4809

Government of Canada, P. S. & P. C. (2007). *Guidance document on statistical methods for environmental toxicity tests. Method Development and Applications Section, Environmental Technology Centre, Environment Canada. En49-7/1-46E-PDF*. Environment Canada. https://publications.gc.ca/site/eng/9.559583/publication.html

Green, J. W., Springer, T. A., & Holbech, H. (2018). *Statistical analysis of ecotoxicity studies*. John Wiley & Sons. https://doi.org/10.1002/9781119488798

Green, J., W., Springer, T., A., Saulnier, A., N., & Swintek, J. (2014). Statistical analysis of histopathological endpoints. *Environmental Toxicology and Chemistry*, *33*, 1108–1116. https://doi.org/10.1002/etc.2530

Hedderich, J., & Sachs, L. (2020). *Angewandte Statistik: Methodensammlung mit R*. Springer. https://doi.org/10.1007/978-3-662-62294-0

Hothorn, L. A., & Kluxen, F. M. (2020). Statistical analysis of no observed effect concentrations or levels in eco-toxicological assays with overdispersed count endpoints. bioRxiv. https://doi.org/10.1101/2020.01.15.907881

Hotopp, I., & Fuelling, O. (2019). *The minimum detectable differences—A way to estimate the power of a small mammals field effects study a posteriori [Poster presentation]*. SETAC Europe 29th Annual Meeting, Helsinki, Finland.

Hotopp, I., Russ, A., Alkassab, A., Pistorius, J., Prados, E., A., & Persigehl, M. (2024). Using equivalence tests in higher tier studies of honey bees under the revised EFSA Bee Guidance -How? *Integrated Environmental Assessment and Management*, *20*, 1496–1503. https://doi.org/10.1002/ieam.4911

International Organization for Standardization (ISO). (2006). *ISO/TS 20281: Water quality—Guidance on statistical interpretation of ecotoxicity data*. International Organization for Standardization.

Jager, T., Albert, C., Preuss, T., G., & Ashauer, R. (2011). General unified threshold model of survival—A toxicokinetic-toxicodynamic framework for ecotoxicology. *Environmental Science & Technology*, *45*, 2529–2540. https://doi.org/10.1021/es103092a

Jager, T., & Ashauer, R. (2025). *Modelling survival under chemical stress: A comprehensive guide to the GUTS framework (Version 3.1)*. Toxicodynamics Ltd. https://leanpub.com/guts_book

Kluxen, F. M., Weber, K., Strupp, C., Jensen, S. M., Hothorn, L. A., Garcin, J.-C., & Hofmann, T. (2021). Using historical control data in bioassays for regulatory toxicology. *Regulatory*

Toxicology and Pharmacology: RTP, 125, 105024 doi:10.1016/j.yrtph.2021.105024PMC: 34364928

Kooijman, S. A. L. M. (2000). *Dynamic energy and mass budgets in biological systems* (2nd ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511565403

Kooijman, S. A. L. M. (2010). *Dynamic energy budget theory for metabolic organisation*. Cambridge University Press. https://doi.org/10.1017/CBO9780511805400

Lehmann, R., Bachmann, J., Karaoglan, B., Lacker, J., Lurman, G., Polleichtner, C., Ratte, H., T., & Ratte, M. (2018). The CPCAT as a novel tool to overcome the shortcomings of NOEC/LOEC statistics in ecotoxicology: A simulation study to evaluate the statistical power. *Environmental Sciences Europe*, 30, 50. https://doi.org/10.1186/s12302-018-0178-5

Lehmann, R., Bachmann, J., Maletzki, D., Polleichtner, C., Ratte, H. T., & Ratte, M. (2016). A new approach to overcome shortcomings with multiple testing of reproduction data in ecotoxicology. *Stochastic Environmental Research and Risk Assessment*, 30, 871–882. https://doi.org/10.1007/s00477-015-1079-4

Mair, M., Kattwinkel, M., Jakoby, M. O., & Hartig, F. (2020). The minimum detectable difference (MDD) concept for establishing trust in nonsignificant results: A critical review. *Environmental Toxicology and Chemistry*, 39, 2109–2123. https://doi.org/10.1002/etc.4847

Nisbet, R. M., Muller, E. B., Lika, K., & Kooijman, S. A. L. M. (2000). From molecules to ecosystems through dynamic energy budget models. *Journal of Animal Ecology*, 69, 913–926. https://doi.org/10.1111/j.1365-2656.2000.00448.x

Organisation for Economic Co-operation and Development (OECD). (1998). *Report of the OECD Workshop on Statistical Analysis of Aquatic Toxicity Test Data*. OECD Publishing, ENV/MC/CHEM(98)18, Paris, France.

Organisation for Economic Co-operation and Development (OECD). (2000). *Test No. 216: Soil microorganisms: Nitrogen transformation test*. OECD Publishing. https://www.oecd.org/en/publications/test-no-216-soil-microorganisms-nitrogen-transformation-test_9789264070226-en.html

Organisation for Economic Co-operation and Development (OECD). (2006). *No. 54: Current approaches in the statistical analysis of ecotoxicity data: A guidance to application*. OECD Publishing. https://www.oecd-ilibrary.org/environment/current-approaches-in-the-statistical-analysis-of-ecotoxicity-data_9789264085275-en

Pires, A., M., Branco, J., A., Picado, A., & Mendonça, E. (2002). Models for the estimation of a "no-effect concentration. *Environmetrics*, 13, 15–27. https://doi.org/10.1002/env.501

Polcher, A., Schlechtriem, C., & Hahn, S. (2023). Review of the OECD Test Guidelines relevant to environmental assessment with regard to the state of the art in science and technology *(Texte 72/2023; Report No. FB001033 ENG, 474 pp.)*. *German Environment Agency (Umweltbundesamt)*, https://www.umweltbundesamt.de/publikationen/review-of-the-oecd-test-guidelines-relevant-to

R Core Team. (2024). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Ritz, C. (2010). Toward a unified approach to dose-response modeling in ecotoxicology. *Environmental Toxicology and Chemistry*, 29, 220–229. https://doi.org/10.1002/etc.7

Ritz, C., Baty, F., Streibig, J. C., & Gerhard, D. (2015). Dose-response analysis using R. *PLoS One*, 10, e0146021. https://doi.org/10.1371/journal.pone.0146021

RIVM—National Institute for Public Health and the Environment (2012). PROAST [Computer software]. https://www.rivm.nl/en/proast. Date retrieved June 20, 2025.

Szöcs, E., & Schäfer, R. B. (2015). Ecotoxicology is not normal: A comparison of statistical approaches for analysis of count and proportion data in ecotoxicology. *Environmental Science and Pollution Research International*, 22, 13990–13999. https://doi.org/10.1007/s11356-015-4579-3

Szöcs, E., & Schäfer, R. B. (2016). Statistical hypothesis testing—To transform or not to transform? *Integrated Environmental Assessment and Management*, 12, 398–400. https://doi.org/10.1002/ieam.1748

U.S. Environmental Protection Agency (U.S. EPA). (2000). BMDS Online (25.1; pybmds 25.1) [Computer software]. https://bmdsonline.epa.gov. Date accessed June 20, 2025.

van Dam, R. A., Harford, A. J., & Warne, M. S. J. (2012). Time to get off the fence: The need for definitive international guidance on statistical analysis of ecotoxicity data. *Integrated Environmental Assessment and Management*, 8, 242–245. https://doi.org/10.1002/ieam.1296

Van Der Hoeven, N. (2004). Current issues in statistics and models for ecotoxicological risk assessment. *Acta Biotheoretica*, 52, 201–217. https://doi.org/10.1023/B:ACBI.0000043442.05899.8c

Vermeiren, P., Wittwer, T., Coffinet, S., von Blanckenhagen, F., & Jakoby, O. (2025). Getting most out of ecotoxicological field data with generalized linear and additive mixed models: A 6-step analysis framework. *Integrated Environmental Assessment and Management*, Forthcoming.

Zuur, A. F., & Ieno, E. N. (2016). A protocol for conducting and presenting results of regression-type analyses. *Methods in Ecology and Evolution*, 7, 636–645. https://doi.org/10.1111/2041-210X.12577

Zuur, A. F., Ieno, E. N., Elphick,., & C. S. (2010). A protocol for data exploration to avoid common statistical problems. *Methods in Ecology and Evolution*, 1, 3–14. https://doi.org/10.1111/j.2041-210X.2009.00001.x

Integrated Environmental Assessment and Management, 2026, 22, 706–716
https://doi.org/10.1093/inteam/vjaf113
Workshop Synthesis