How we can make ecotoxicology more valuable to environmental protection

Hanson, M. L., Wolff, B. A., Green, J. W., Kivi, M., Panter, G. H., Warne, M. S. J., Ågerstrand, M., Sumpter, J. P.

Original citation & hyperlink:
Hanson, ML, Wolff, BA, Green, JW, Kivi, M, Panter, GH, Warne, MSJ, Ågerstrand, M & Sumpter, JP 2016, 'How we can make ecotoxicology more valuable to environmental protection' Science of The Total Environment, vol (in press). DOI: 10.1016/j.scitotenv.2016.07.160
https://dx.doi.org/10.1016/j.scitotenv.2016.07.160

DOI 10.1016/j.scitotenv.2016.07.160
ISSN 0048-9697
ESSN 1879-1026

Publisher: Elsevier

How we can make ecotoxicology more valuable to environmental protection?

Hanson ML[1*], Wolff BA[2], Green JW[3], Kivi M[4], Panter GH[5], Warne MStJ[6,7], Ågerstrand M[8], Sumpter JP[9]

[1]Department of Environment and Geography, University of Manitoba, Winnipeg, Canada

[2]Department of Fish, Wildlife and Conservation Biology, Colorado State University, Fort Collins, Colorado, United States of America

[3]DuPont, Wilmington, Delaware, United States of America

[4]Health Canada, Ottawa, Canada

[5]Syngenta International Research Centre, Jealott's Hill, United Kingdom

[6]Queensland Department of Science, Information Technology and Innovation, Brisbane, Queensland, Australia

[7]Centre for Agroecology, Water and Resilience, Coventry University, Coventry, United Kingdom

[8]Department of Environmental Science and Analytical Chemistry (ACES), Stockholm University, Stockholm, Sweden

[9]Institute for the Environment, Health and Societies, Brunel University London, United Kingdom

*Contact Author Email: mark.hanson@umanitoba.ca

ABSTRACT

There is increasing awareness that the value of peer-reviewed scientific literature is not consistent, resulting in a growing desire to improve the practice and reporting of studies. This is especially important in the field of ecotoxicology, where regulatory decisions can be partly based on data from the peer-reviewed literature, with wide-reaching implications for environmental protection. Our objective is to improve the reporting of ecotoxicology studies so that they can be appropriately utilized in a fair and transparent fashion, based on their reliability and relevance. We propose a series of nine reporting requirements, followed by a set of recommendations for adoption by the ecotoxicology community. These reporting requirements will provide clarity on the experimental design and conditions, chemical identification, test organisms, exposure confirmation, measurable endpoints, how data are presented, data availability and statistical analysis. Providing these specific details will allow for a more full assessment of the reliability and relevance of the studies, including limitations. Recommendations for the implementation of these reporting requirements are provided herein for practitioners, journals, reviewers, regulators, stakeholders, funders, and professional societies. If applied, our recommendations will improve the quality of ecotoxicology studies and their value to environmental protection.

Keywords – publications, quality, reliability, relevance, risk assessment, reporting recommendations, peer review

Introduction

There is widespread and growing concern that the quality, usability, and reporting of published peer-reviewed research is not as good as it could, and should, be. This can undermine the credibility and functioning of the scientific endeavor (Alberts et al., 2014; Forbes et al., 2016) and is a conversation that has spread beyond just the scientific community (e.g. The Economist, 2013). Poor science and reporting also come with steep economic costs. For example, it has been estimated that irreproducible results in the biomedical literature cost 28 billion USD in America alone, each year (Freedman et al., 2015). In addition to the direct economic costs from repeating poorly conducted studies that report spurious results, poor quality research delays and hinders protection of the environment, which is an underlying reason for conducting ecotoxicology research. By performing and reporting poor science, as a discipline we are failing to achieve our existential goal.

Only recently has attention been focused on the quality of published ecotoxicology studies (e.g. Klimisch et al. 1997; Durda and Preziosi 2000; Hobbs et al. 2005; Schneider et al. 2009; Brady, 2011; Agerstrand et al. 2014; Warne et al. 2015). A core issue is that ecotoxicology studies often do not report key information required to make a judgment on the quality of the studies (Harris and Sumpter, 2015). Harris et al. (2014) suggested twelve principles that they considered should be addressed in an ecotoxicology study. They then applied three of the most objective of these principles (i.e.. measurement of exposure

concentrations, study repeated, and more than a single exposure tested) to 200 randomly chosen research papers, published in 2013, in three reputable journals covering the field. They concluded the quality of published ecotoxicology research was poor, with less than half, and often less than 25%, fulfilling the criteria. In particular, often less than 25%of papers provided information demonstrating that results were repeatable, with the majority of papers reporting results from only one experiment.

To further evaluate the quality of published ecotoxicology studies, we objectively assessed current reporting requirements of journals publishing ecotoxicology studies. This involved conducting an ISI Web of Science search using the topic keyword ‘ecotoxicology’, and years published ‘2014 – 2015’. The initial search generated 176 journals that published ‘ecotoxicology’ studies. We then employed a cut-off requiring greater than two ‘ecotoxicology’ articles published in 2014 – 2015. This cut-off left 31 journals, and for those we then screened the ‘guide to authors’ for three basic criteria we deemed fundamental to paper quality in ecotoxicology. These criteria were: 1) expectations around statistical analysis (e.g., adequate replication); 2) analytical verification of exposure concentrations; and 3) availability of Supplemental Information (as a mechanism of providing data and other information required for critical analysis of relevance and reliability). We found that relatively few journals provided guidance on our three criteria, and only one journal met all three (Figure 1). This exercise suggests that guidance on publication standards provided by peer-reviewed journals requires

improvement. We do acknowledge that many journals rely on reviewers to assess publication quality, including use of our search criteria (i.e., analytical confirmation of exposure concentrations, statistical guidance). However, relying on expert judgment alone can be problematic, as it is often inconsistently applied between reviewers (e.g., depends on the reviewers' expertise and availability, as well as their own biases (Mahoney, 1997)) and is coupled with the needs of a journal to fill issues and increase impact factors. Baseline expectations around reporting and conductance of studies by all journals, along with mechanisms to facilitate a full review by future readers (e.g., through availability of Supplemental Information), are required in order to elevate the quality of the ecotoxicology literature as a whole (Moermond et al. 2016).

Overall, the reporting of both the methodology and results in ecotoxicity studies appears to be incomplete and inadequate (Ågerstrand et al., 2011a, 2011b, 2014). This can decrease the likelihood that studies are cited by other authors, or used for regulatory purposes (ECHA, 2012). Examples of missing or insufficiently reported aspects include the types and performance of controls, analytical methods and exposure confirmation, test system design, information about statistical evaluations and statistical power, and presence of possible confounding factors. As a reader of a peer-reviewed publication it can be challenging to decide whether missing information is due to insufficient reporting or inadequate design and performance of the experiment. Regardless of the

cause, missing information decreases the value of ecotoxicology publications and may lead to a paper being omitted from subsequent interpretative work.

Only when studies are reported in a transparent and detailed way is it possible for a reader (e.g. a regulator who wants to use the results in the publication to ensure adequate protection of the environment) to judge the quality/reliability of the paper. Until recently, word limits of peer-reviewed journals caused authors to economize on the details of methods and materials, in order to leave room for descriptions of the results and discussion. This is no longer necessary, since it is possible to publish supporting information online where all aspects of a study can be reported in sufficient detail and raw data provided (e.g. Meyer and Francisco, 2013). In this work we recommend a set of minimal reporting requirements. We then suggest a variety of strategies to maximize uptake of these requirements into all published ecotoxicology papers by those most able to create the necessary change within the peer review process (e.g., authors, journals and peer reviewers).

RELEVANT REPORTING REQUIREMENTS FOR ECOTOXICITY STUDIES

To ensure detailed and transparent reporting of peer-reviewed publications, reporting recommendations can be used as a tool for designing, performing, analyzing and reporting ecotoxicity studies (Ågerstrand et al., 2014; Moermond et al. 2016). This has also been suggested in other areas, e.g. in biomedical

research, to increase the value of publications and reduce waste from the reporting of poorly written research articles (Glasziou et al., 2014, Chan et al., 2014). Using a systematic reporting tool also has the potential to shorten time needed for peer-review and to decrease the number of questions from peer-reviewers, thereby increasing the chance of a paper getting published. It is a great loss in terms of scientific knowledge, but also from the viewpoint of animal welfare and economic resources, when peer-reviewed literature cannot be used in hazard and risk assessment of chemicals. Below we provide, in no particular order of importance, nine general reporting requirements to help enhance the quality, credibility, and usability of the ecotoxicology literature and a checklist for both authors and peer reviewers to employ when writing and assessing ecotoxicology studies (see and example in Table 1).

1. Test Compound Source and Properties

Test substances or products may have more than one component contributing to their toxicity and varying the amounts of these components can affect toxicity. As manufacturing processes change over time, the composition of a substance and its toxicity may also change. Ecotoxicology data derived from a historical form of a test substance may not be relevant to current forms of the test substance. As an example, the US EPA required an upper limit of 0.1% be established for DDT and its related impurities (ΣDDT, i.e. DDT, DDE and DDD) for all dicofol technical active ingredients by 1987 (US EPA, 1998). Unless otherwise specified, any data

produced prior to this date could have been conducted with technical active ingredients containing greater levels of contamination and therefore inaccurate estimates of the biological properties of the current product.

In another example, the pesticide cyfluthrin contains eight isomers, of which four are more biologically active. The related pesticide, beta-cyfluthrin, only contains the four more biologically active isomers. When the isomer profile is known, the exposure and ecotoxicity of both substances can be compared by correcting for isomer-equivalents, as was done by the US EPA for the aquatic risk assessment for cyfluthrin and beta-cyfluthrin (US EPA, 2013).

The following are the minimum recommended reporting requirements for test substance identification, where available:

- Technical name (e.g. International Union of Pure and Applied Chemistry (IUPAC) or registration no., Chemical Abstract Service number (CAS), batch number) and formulated product, brand or trade names;
- Source, purity, and composition of the test substance; specifically, percent active ingredient including levels or ratios of components and isomers and impurities

Basic physico-chemical property information (e.g., aqueous solubility, acid dissociation constant (pKa), dissociation constant (Kd), octanol-water partition coefficient (Kow), organic carbon-water partition coefficient (Koc), vapour

pressure (VP) or Henry's Law Constant (HLC), Bioconcentration Factor (BCF), Bioaccumulation Factor (BAF) and Biomagnification Factor (BMF)) can ensure studies are designed to minimize losses (e.g., volatilization, degradation or adsorption to test vessels; see Section 5 Exposure Confirmation) as well as identify the potential for chemical reactions to occur under different test conditions (e.g., ionization in relation to pH; see Section 4 Experimental Conditions). For example it might be relevant to test:

- a substance with an environmentally relevant pKa at various pHs to account for any differential uptake between its neutral and ionized form.
- a rapidly degrading, volatilizing, or strongly partitioning substance in a flow-through test system to ensure constant exposure concentrations.

2. Experimental Design

Details of the experimental design should be stated or, if the design is complex, a figure can often more efficiently and accurately explain the design (e.g. Figure 2 from Dellinger et al., 2014). It is important that the experimental design should permit the hypothesis to be tested and the objectives to be met. This can include incorporating test conditions and species reflective of where the data will be applied by regulators (See Section 3: Test Organism Characteristics and Section 4: Experimental Conditions). For example, the European Food Safety Authority (EFSA 2013) requires that the test conditions and species are reflective of the regions undergoing assessment, and this lack of congruency (particularly in soil

tests) has been a major factor in its rejecting certain studies in its evaluations.

Experimental design features that should be reported include:

- Hypotheses and objectives of the study should be clearly stated, even if the hypothesis is as simple as ‘compound X causes a 50% reduction in egg-laying relative to control at a concentration less than its aqueous solubility’.
- The number of treatments and the nominal concentrations;
- The number and types of controls (e.g., positive, negative, or solvent). If a solvent carrier is used then its concentration in each treatment and control should be equal and stated.
- The degree of replication of each treatment and control and an explanation of whether they are true replicates or pseudo-replicates.
- The methods for creating and storing the stock and working solutions and the duration of storage.
- The exposure regime for the test substance (e.g. static, semi-static or flow-through) with details about the renewal regime and method.
- The method/design for determining the order in which test organisms are added to test vessels and the placement of test vessels (e.g. randomized, stratified random or Latin square design).
- The frequency of exposure (e.g., how often the test substance is administered or renewed) and type of samples analyzed to determine toxicant concentrations.

- Route of exposure to test organism should be clearly stated (e.g. via the diet, via the media, etc.).
- Details of all quality assurance and quality control procedures conducted as part of the study (e.g. whether the design was blind or double blind, whether scoring, data entry and calculations were conducted by one, or more than one individual, and see Section 4: Experimental Conditions as they relate to water quality, etc.).

3. Test Organism Characteristics

Test organisms may be obtained from a variety of sources, including in-house cultures or wild populations. Providing details on where they were obtained, how they were maintained, and as much information as possible on the control performance of the test species, is useful when determining if a chemical, or mixture of chemicals causes an effect and if that effect can readily and reliably be detected. The following points should be considered and reported, where applicable:

- Species selection – Justify the selected organism (e.g. ecological relevance and/or relevance to hypothesis)
- Identity of the species – Report the common and scientific name of the species, general type (e.g., plant), its source, and strain (if appropriate). Provide the DNA Bar-Code details if available. When dealing with new or cryptic species, genetic identification is recommended, e.g., the alga

Oophila sp. (Baxter et al., 2015) or the *Hyalella azteca* species complex (Leung et al., 2016).

- Source – In-house cultures, wild populations and, where applicable, to include method of collection (e.g., collection of fertilized eggs, or animals from the wild) and their subsequent handling.
- Life-history – The stage of the life-cycle, the age and sex of the test organisms, and their size or mass at the beginning of the experiment, should be provided.
- Husbandry – All procedures related to maintaining the organisms in good condition should be stated. Overviews of welfare and ethical approvals need to be reported, especially those that may influence observed responses (e.g., degree of enrichment, groupings). Outbreaks of disease or unexplained morality/morbidity, including their incidence and severity, and how these were treated, must be reported.
- Test species performance – Available historical data on endpoints (e.g., growth rates, reproduction) used in the experiment should be provided, enabling the data collected during the experiment to be put into context. For example, a detectable change in growth may not be deemed biologically significant when compared to historical control data of the performing laboratory (Länge et al., 2001). Control performance should be reported in order to permit comparison to validation criteria.

4. Experimental Conditions

In addition to experimental design (See Section 2: Experimental Design) the general conditions of the experiment, facilities, and operating regime should be provided, as the interaction of the test organism and testing/exposure environment influence the outcome of the study. Where applicable, the following should be reported:

- General testing facilities – For example, growth chamber (make, model), tank (dimensions, capacity, rate of water change); mesocosm (location, volume, flow rate), greenhouse (location, size), field location (GPS coordinates, general climatic/environmental information, for duration of study).
- Test conditions – All available details on relevant experimental parameters such as light intensity, photoperiod, and temperature should be reported as means, with the variability.
- Source, type and composition of test media – e.g. water, commercially available media, soil, sediments, including known background contaminants.
- Test media parameters – All measurements that can influence test organism health or change endpoint responses should be reported (e.g. dissolved oxygen concentration, temperature, pH, salinity) as means with an estimate of variability (i.e., confidence intervals). In addition, report the properties of the test media that may influence interpretation of the test results (e.g. dissolved organic carbon where binding of test material is expected).

- Dosing mechanism – e.g. via the diet, peristaltic pump, spray application.
- Details of acclimation – For both the test system and test organisms. In terms of the test system, this is to demonstrate that the conditions are stable prior to introduction of test organisms. This is of particular importance in mesocosm and sediment studies. In terms of the test organisms, this is to ensure survival and growth of test organisms during the experimental conditions (further details see Section 3 Test Organism Characteristics).
- Feeding – Information should include the type of food, source, amount provided and frequency of feeding. In the case of commercial foods, detailed reporting of characteristics as required. The concentration of any contaminants should be included, where relevant.
- Number and density of organisms – This may be influenced by purpose of the test or experimental design (e.g., test power, see Section 8: Statistical Analysis). However, test design should enable normal behavior of the test organism. Density of the test organisms will determine adequate feeding requirements and acceptable loading.
- Good Laboratory Practices (GLP) studies – When data are from a GLP study, this should be expressly stated in the publication, as well as the location of the raw data.
- Quality Assurance and Quality Control (QA/QC) activities – (e.g. calibration of laboratory equipment) and the results of these should be

reported. If Standard Operating Procedures (SOPs) are used provide location of these.

5. Exposure Confirmation

Characterization of exposure in an ecotoxicity test is critical to facilitating publication, ensuring the stated hypothesis is being addressed, reducing uncertainty in the observed relationships, and allowing risk assessors to incorporate the data into their evaluations. Strong analytical support provides confidence that the stated chemical was the one that was used, that it was found in the test vessels at the concentrations targeted, in the appropriate compartments, with an understanding of the true duration of exposure. There are numerous examples in the scientific literature where the lack of analytical confirmation has resulted in erroneous conclusions, costly follow-up work, and retractions of published works (e.g., Ricaurte, 2003).

What follows are minimum reporting requirements around exposure confirmation for conducting standard laboratory ecotoxicology tests, but they can also be applied to *in vitro*, micro- and mesocosm, and field studies.

- Provide sufficient instrumental and details around your analytical approach, or a suitable reference, that supports the approach taken.
- Report any relevant QA/QC undertaken during the sampling and analysis (e.g., blanks, storage studies, internal standards, recovery efficiency, and specific storage preservation techniques) and the results of the QA/QC.

- Report your limits of detection, quantification, or reporting (LOD, LOQ, and LOR, respectively) and their variance.
- State clearly what samples were analyzed (e.g., stocks only, exposure vessels, pooled or unpooled) and the timing or frequency of measurements. This is important when interpreting the relevance of the observed response in light of actual exposure duration, regardless of test length.
- State the media type and the volumes sampled, as well as storage conditions and time till analysis.
- Report your values in metric units as the target analyte, and not as the formulated product. Where applicable, provide means of measured values and standard deviations/errors.
- State clearly whether subsequent statistical analyses and interpretations rely on nominal or measured values (See Section 8), and whether these exposures values have been corrected for recovery.
- Prepare a plan for archiving your original data (See Section 9).

There are some scenarios where robust analytical support is not feasible or for which analytical methods below the required reporting requirements are unlikely to be met. Still, in the case of substances where routine measurements and protocols for analysis do exist, as well as very cost effective and relatively simple analytical approaches (e.g., enzyme-linked immunosorbent assay; ELISA), no standard toxicity test should be performed without some exposure confirmation. If

366 you are unable to provide a level of analytical support that gives risk assessor's
367 confidence in the data, this will seriously reflect on the value of conducting the
368 study.
369
370 **6. Endpoints**
371
372 The reporting of information around test endpoints is especially important when
373 assessing the relevance and reliability of the data. While endpoints, such as
374 mortality, reproduction and growth, are familiar to the majority of ecotoxicologists,
375 there are many that are less familiar (e.g., genomic and metabolic tools). This
376 can lead to misunderstandings and misinterpretations of the significance and
377 ecological relevance of the data that can and should be avoided. What follows
378 are reporting recommendations around endpoints in ecotoxicology tests:
379 - State all endpoints monitored in the study, regardless of the observed
380 response (e.g., avoid reporting only 'differences')
381 - Define the endpoint in order to remove ambiguity (e.g., what is a
382 'malformation'?)
383 - Justify the selection of your endpoints (also see Section 2 Experimental
384 Design) and their statistical power (see Section 8 Statistical Analysis).
385 - Express clearly when and how the endpoint was monitored and recorded
386 (e.g., blind evaluations of behavior; See Section 2 Experimental Design)
387 and how these data are presented in the paper (e.g., Tables,
388 supplemental information)

- Report other observations that may have relevance, but were not an explicit part of the original study design (e.g., lesions in fish), as this can inform future work and be hypothesis generating.

7. Presentation of Results and Data

Presentation of results is an important aspect of ecotoxicology that is often not discussed but has implications for assessing the utility of peer-reviewed literature. The primary purpose of Figures and Tables is to convey as much information as possible, in a manner that is simple to understand. For excellent examples of good graphics, see Tufte (1997, 2001). It is equally important to report figures that allow an assessment of the statistical interpretation and inference. To facilitate this we recommend:

- Create figures that provide readers with greater ability to assess data distributions and variability (e.g., scatterplots, histograms, box plots, etc.) for each time-point. It is common for researchers to rely upon graphs displaying a mean ± standard deviation or standard error. However, this has been shown to be problematic because different distributions of data can be represented in the exact same way when relying solely upon bar charts and line charts (Weissgerber et al. 2015). Furthermore, traditional bar charts can disguise outliers in data, which can be important, particularly for studies with small sample sizes (Weissgerber et al. 2015).

- Employ appropriate scales (e.g., do not truncate or break axes to over-emphasize effect sizes or differences relative to controls).
- Provide details in figure or captions specifying the statistical test employed (if applicable), degree of replication, level of statistical significance (if any).
- Confidence intervals, with alpha-level, should accompany summary statistics in figures and tables, instead of standard deviation and standard error. Confidence intervals are preferred over standard deviations or standard errors because measures of uncertainty are not always symmetric about the mean. This is especially relevant when the data are transformed for analysis and the results are expressed in back-transformed values. It also applies when some non-normal error structures are used, such as Poisson or binomial. Only in the case of normally distributed data, will standard deviations provide equivalent information to confidence intervals.
- Inclusion of statistically and biologically significant, as well as non-significant results, will allow for a balanced understanding of the full range of responses.
- It is useful to report summary data (e.g., end-points, estimates, ranges, etc.) in table-format, and greater attempts should be made to include as much data as practical, e.g., in Supplemental Information (See Section 9 Raw Data).

- For field studies it is helpful to provide maps with relevant information (e.g., GPS coordinates, scale bars, orientation, regional context, etc.) that may enable a reader to understand the spatial context of a study.

8. Statistical Analysis

A well-conducted experiment can have its meaning distorted if poor statistical methods are used in the interpretation. Consequently, care should be taken in the selection and interpretation of statistical tests and models, and subsequent reporting of the approaches employed. Considerations include the following:

- Provide a statistical flow chart that includes any preliminary data checks to satisfy test or model requirements and accommodations (e.g., transformations or robust methods) of data problems, the handling of multiple controls (e.g., solvent and negative controls), and indicate how statistical tests or models are selected and why (e.g., OECD 2006, 2010).
- Provide justification for any transformations (e.g., logarithm) used and whether/how it affects the analysis results.
- Power of the planned hypothesis testing procedure to find effects or the ability of a regression model to estimate ECx reliably should be reported (See Section 4: Experimental Conditions). This can be done partly through the use of historical control data (See Section 3 Test Organism Characteristics).

- Good estimation of the control mean is important since all tests and estimates are in relation to that mean, so if there were more replicates in the control than in treatment groups or other special considerations of the control, make clear what was done and why.
- Statistical outliers should be identified and their effect on conclusions should be stated.
- In reporting sub-lethal effects in a study with substantial mortality in high treatment groups or loss of subjects/replicates for other reasons, report any adjustments made to the tests or models (e.g., weighting) to avoid over-interpretation resulting from small sample size.
- Explain how the statistics account for the actual experimental set-up (e.g., individual, paired or group housing, expected monotone dose-response or deviations therefrom, such as hormesis)
- For complex models (for regression or hypothesis testing) with multiple potential explanatory variables, any model selection method that sequentially adds or removes terms to arrive at a final model should be described and, if possible, verified by alternative approaches to avoid unintentional bias.
- Results should be reported in the original units and with no more significant digits than the raw data justify. For example, if the data are measured with two significant digits to the right of the decimal point, it is pointless to report means to five decimals points. That implies a level of precision not justified by the data. Moreover, the quality of the

measurements determines how many significant digits are meaningful. If the equipment being used to measure a response is accurate only to the nearest 0.1, but reports five digits past the decimal point, data should be reported only to the nearest 0.1. Summary statistics should be reported as means and confidence intervals (not standard deviations or standard errors) with an explanation of how confidence intervals were determined. This is especially important if transformed data were analyzed, so confidence intervals are not symmetric about the mean.

- While expressing change from control (percent of control) is useful for presentation, data should also be presented as actual recorded values.

There is disagreement about the appropriateness of calculating and implementing certain measures of toxicity, specifically no observed effect concentrations (NOECs) and lowest observed effect concentrations (LOECs) (Green et al., 2012; van Dam et al., 2012 and references therein). Still, they are necessary for some responses and datasets and are commonly used in many risk assessment frameworks, and so we have provided specific reporting requirements below.

- Power to detect a specified size effect should be stated.
- Confidence intervals for the mean response at the NOEC and LOEC should be reported along with the percent change from control.
- If the response is non-monotonic (e.g., hormesis), indicate how this was addressed in the statistical analysis.

Effective concentrations (ECx point estimates where x is typically no more than 50) also have specific reporting requirements. These are:

- Define percent change from control (the 'x' in ECx) as to whether it applies to raw or transformed data.
- Report when ECx estimates are extrapolations, including extrapolations below the lowest tested positive concentration/dose.
- Report what exposure values (e.g., measured versus nominal) were used in your estimates
- Model selection and goodness-of-fit criteria should be specified.
- The minimum effect size that a regression model can reliably estimate should be determined and reported. For example, if the standard error of the control mean is 20% of that mean, then the estimation of ECx for x <20 is likely unjustified.
- Confidence intervals should be reported for ECx and all model parameters and explain why any parameters not significantly different from zero were included in the model, as these can indicate possible model problems.

9. Raw Data

Ideally, all data that are important to the determination of a toxicity value should be archived in a manner that they can later be made available for validation

and/or re-evaluation. However, the minimum data that should be archived (for example, in the Supplementary Information/material section of journals) are:

- The replicate and treatment identification.
- The measured and nominal chemical concentrations for each analysis.
- The non-transformed (e.g., non-logged, non-normalised) biological effects data at the level of the unit of measurement. For example, data on individuals if that is what is measured or data for groups of individuals when pooled and then measured.
- All measured values for experimental conditions that are known to affect the toxicity or bioavailability of the chemical.

The reason for archiving the above data is to increase the usability and longevity of impact. For example, risk assessors may require the raw data of papers that are crucial to their ecological risk assessments or regulatory decision-making. They may require these data in order to re-analyse them or to confirm the estimate of toxicity, especially as risk protection goals change through time, and across jurisdictions.

WHAT CAN DIFFERENT STAKEHOLDERS DO TO IMPROVE THE REPORTING AND IMPACT OF OUR SCIENCE?

All of us who produce, review, publish and use ecotoxicology data play a vital role in improving the quality and reporting of our science. There are also

545 immense benefits to everyone involved in this endeavor should we get it right.
546 There are a number of actions we can take now to move the discipline forward so
547 that our shared goal of environmental protection is accomplished.
548
549 **A. What Journals Can Do:**
550 Some of the benefits to journals from better ecotoxicology studies would be faster
551 peer reviews, reduced likelihood of damaging retractions, and increased impact
552 factors. By demonstrating a commitment to the best ecotoxicology, journals can
553 distinguish themselves from predatory publishers (Bohannon; 2013; Kolata,
554 2013). Below, we provide advice and guidance to journal publishers, editors-in-
555 chief, associate editors, reviewers and authors.

556 1. Journals should work with all stakeholders to create clear minimal
557 standards for publishing ecotoxicology studies in concert with the peer
558 review process. This could be implemented through a formal checklist for
559 authors prior to submission. Ideally, all journals would implement the same
560 standards. Journals could screen submissions using their checklist and if
561 their criteria are not met, the paper will not be reviewed.
562 2. Reporting author contributions, source of funding, and other possible
563 conflicts of interests, with no addition to author list after submission.
564 3. Journal training of reviewers and the use of the checklist of publishing
565 criteria (same checklist as journal submission) to facilitate review.
566 4. Open discussion/critique of papers and mechanisms for discussion, such
567 as letters to the editor in online forums.

5. Create more effective mechanisms for corrections by authors, and clear reporting of retractions, with the reasons, for published papers.
6. Facilitate obtaining data reported in papers (e.g., supplemental information, figures with extractable data). Supplemental Information should download with the PDF of the paper (e.g., as it does with Proceedings of the National Academy of Sciences, USA).
7. Encourage submission of good quality papers that contain negative findings, or that replicate (or fail to) previous studies.

B. What Scientists/Practitioners Can Do:

The benefits of improving the conductance and reporting by scientists (whether academic, government, industry, consultants and contract labs, or non-government organizations (NGOs), and students) include less time, resources and money wasted repeating previous work that was poorly performed by others, fewer animals being used, studies being reviewed more rapidly and by better journals, greater inclusion of publications in decision-making processes. Articles with greater impact can also lead to increased funding to do more good science. To achieve these goals we recommend:

1. Work towards enhancing the training of all practitioners prior to the conduction of any ecotoxicology study.
2. Prior to starting the study, draw in appropriate expertise to ensure greatest possible quality (e.g., statisticians, chemists).

3. Create a checklist of a good quality study prior to commencing experimental work (see Section A: What Journals Can Do; Example in Table 1) and have a plan to meet those requirements.
4. Develop internal laboratory QA/QC procedures and training.
5. Attempt to verify unusual results (e.g., replicate study).
6. Acknowledge the limitations of your data and do not over-interpret the results.
7. Cite good quality and appropriate science. As a general rule of thumb, do not cite retracted papers.
8. Have a mechanism for storing data (historical assay performance, etc.) and a means to share that with others. For example, graduate students can include raw data within their thesis appendices.
9. Push for a set of consistent screening toxicity test methods across jurisdictions and standard test organisms so that replication is facilitated and these lower tier data have the widest possible uptake by regulators. With this recommendation, we are not attempting to stifle scientific inquiry, but to reduce the ambiguity that can arise when different investigators ask the same initial screening question, e.g., what is the response of the duckweed *Lemna minor* when exposed to a chemical?
10. Become familiar with testing standards standards (e.g., Japanese Ministry of Agriculture, Forestry and Fisheries (JMAFF), Office of Chemical Safety and Pollution Prevention (OCSPP), Organisation for Economic Co-operation and Development (OECD) and the American Society for Testing

and Materials (ASTM))for your test organisms, so that you are aware of performance requirements and minimal expectations around experimental design.

11. Investigate whether or not regulatory bodies have criteria for evaluating published literature information (e.g., ECHA 2012, EFSA 2013, US EPA 2011) and report this information in your study.
12. Practitioners should gain a better understanding of GLP and its role in improving the quality of reports for risk assessment (see Borgert et al., 2016).
13. Encourage conversations and consultation with diverse data users (e.g. NGOs, media, industry).

C. What Data Users Can Do:

The benefits to data users from the better reporting of ecotoxicology includes the context required to interpret and integrate the results with all other information evaluated during a risk assessment, less uncertainty in the decision-making process, studies and data that can be used across regulatory jurisdictions, saving time and money, and positions based on data that will be more broadly accepted. It is acknowledged, however, that given the precautionary approach taken in most risk assessment frameworks, even if a study does not report all experimental details, it may still be considered if it provides information on a potential relevant risk not assessed by other data (e.g., toxicity to a non-standard sensitive species/population or vulnerable ecosystem). Those who read and use

636 the ecotoxicology literature, such as risk assessors, regulators, NGOs, policy
637 makers, risk managers, scientists/practitioners play an integral role in improving
638 the quality and reporting of ecotoxicology. This can occur through:

639 1. Setting and making available clear expectations for well-conducted
640 studies, guidance to assess those studies, and the outcomes of those
641 assessments.
642 2. Engaging with the totality of the data and justify the inclusion/exclusion of
643 studies and explain how information from multiple studies is assessed,
644 weighted and integrated together.
645 3. Consistently use and cite the best science available.
646 4. Conduct outreach amongst practitioners and data users to communicate
647 what your data needs are and why.
648 5. Work towards setting consistent toxicity test methods across jurisdictions.

649

650 **D. What Other Stakeholders Can Do:**

651 In this instance, we are thinking primarily of the public and media, and what they
652 can do to help ensure better ecotoxicology. Some simple things would be to:

653 1. Cite good quality and appropriate science in an unbiased fashion.
654 2. Encourage consultation with the range of scientists/practitioners and data
655 users (academics, government, industry).
656 3. Work towards building a stronger scientific understanding of what the
657 strengths and limitations of peer reviewed literature in general are,
658 become comfortable with uncertainty, and acknowledge that well-

conducted and reported studies will trump poorly reported and conducted studies.

E. What Funders Can Do:

We acknowledge that the vast majority of the ecotoxicology that is done is a result of funding, whether from government, industry, or other sources, such as NGOs. As bodies that decide which work will be performed, it is vital they ensure that they strive to support the highest quality science, and that it is reported properly. The benefit to funders will be the creation of data that will allow for the widest reach by all users, enhancing the value of limited financial resources. To facilitate this, funders should:

1. Create minimum requirements for conducting studies and reporting of data prior to funding approval.
2. Provide funding so that open source publishing and repositories for raw data can be maintained.
3. Promote ethics and integrity among grantees.

F. What Professional Societies Can Do:

Many of us involved in the field of ecotoxicology are also members of professional societies (e.g., Society of Environmental Toxicology and Chemistry) that can bring together ecotoxicologists from all stakeholder groups, as well as provide forums such as dedicated journals for the dissemination of new data.

Through these societies we can promote better conductance and reporting of ecotoxicology studies through:

1. Better and ongoing training of scientists/practitioners (e.g., free short courses on best practices for the conducting and reporting of studies at annual meetings).
2. Promoting ethics and integrity for students and supervisors in their research activities.
3. Advocating for a set of consistent toxicity test methods across jurisdictions that will be the agreed initial screen characterization of the toxicity of a compound to a particular organism.
4. Promoting civil and open discussion/critique of papers and mechanisms for discussion, such as special sessions at annual meetings.
5. Ensuring society journals are working with publishers, authors, and reviewers to improve the reporting of new and negative data.

DISCUSSION

We believe that our recommended reporting requirements (which also inform practice), coupled with our recommendations to promote communication among users, will improve the overall quality of ecotoxicology. We acknowledge that our recommendations do not begin to address adequately the issue of relevance for the studies themselves (i.e., asking the right question). However, when a user of ecotoxicology data has identified a series of studies as highly relevant, our

recommendations should help him or her to distinguish those that are of the greatest quality and how well they have been performed to address the question of interest (i.e., what is the reliability (aka quality) of the data). We also acknowledge that the requirements we propose are neither definitive nor fixed. As the types of studies we conduct change (e.g., new protocols, new classes of chemicals of concern, *in silico* methods) the reporting requirements might change as well. Finally, we acknowledge the need in some cases for exclusivity of data and feel this is a question of striking the right balance. An appropriate mechanism (e.g., registration protections without the need to keep data from competitors or designating a neutral third party to handle any sensitive information) for sharing can be created so that proprietary data generated for risk assessment and regulators can be examined by all stakeholders. Regulatory agencies are tasked with making scientifically-informed decisions on behalf of the public, and therefore need to use and be seen using data of the highest quality, but also communicating why those data are selected, to ensure public trust and reduce perceptions of possible bias (Forbes et al., 2016).

In summary, we have identified crucial areas where the quality of research and publication can be strengthened. These have been addressed through a set of broad recommendations for everyone involved in the discipline. If these are applied, ecotoxicology and its application in environmental protection will improve.

Acknowledgements – The authors would like to that thank the Society of Environmental Toxicology and Chemistry (SETAC) for organizing the PellstonTM workshop "Improving the usability of ecotoxicology in regulatory decision-making" in Shepherdstown, West Virginia August 30th to September 4th 2015. Support for the workshop was provided by USEPA, CropLife America, Ecetoc, European Crop Protection, Monsanto, Compliance Services International, FMC, Syngenta, BASF, Exponent, ISK Biosciences, Dupont, MERA, Bayer CropScience, Dow AgroSciences, Intrinsik, and SETAC. The commitment of the sponsors to advancing environmental science is appreciated. The sponsors did not influence the content of this paper.

REFERENCES

Ågerstrand M, Breitholtz M, Rudén C. 2011a. Comparison of four different methods for reliability evaluation of ecotoxicity data: a case study of non-standard test data used in environmental risk assessments of pharmaceutical substances. Environ Sci Eur. 223(1):17.

Ågerstrand M, Küster A, Bachmann J, Breitholtz M, Ebert I, Rechenberg B, Rudén C. 2011b. Reporting and evaluation criteria as means towards a transparent use of ecotoxicity data for environmental risk assessment of pharmaceuticals. Environ Pollut. 159(10):2487–92.

Ågerstrand M, Edvardsson L, Rudén C. Bad Reporting or Bad Science? Systematic Data Evaluation as a Means to Improve the Use of Peer-Reviewed Studies in Risk Assessments of Chemicals. 2014. Hum Ecol Risk Assess Int J. 20(6), 1427-1445.

Alberts B, Kirschner MW, Tilghman S, Varmus H. 2014. Rescuing US biomedical research from its systemic flaws. Pro Nat Acad of Sci. 111(16), 5773-5777.

Baxter LR, Brain RA, Hosmer A, Nema M, Muller K, Solomon KR, Hanson ML. 2015. Effects of atrazine on egg masses of the yellow-spotted salamander (*Ambystoma maculatum*) and its endosymbiotic alga (*Oophila amblystomatis*). Environ Pollut. 206:324-321.

763 Bohannon, J. 2013. Who's afraid of peer review. Science, *342*(6154).
764
765 Borgert C, Becker R, Carlton B, Hanson M, Kwiatkowski P, Marty M, McCarty L,
766 Quill T, Solomon K, Van Der Kraak G, Witorsch R, Yi K. Does GLP Enhance the
767 Quality of Toxicological Evidence for Regulatory Decisions? Tox Sci. 151:206-
768 213.
769
770 Brady, D. 2011. Evaluation Guidelines for Ecological Toxicity Data in the Open
771 Literature. USEPA**.** www.epa.gov/pesticide-science-and-assessing-pesticide-
772 risks/evaluation-guidelines-ecological-toxicity-data-open
773
774 Chan A-W, Song F, Vickers A, Jefferson T, Dickersin K, Gøtzche PC, Krumholz
775 HM, Ghersi D, van der Worp HB. 2014. Increasing value and reducing waste:
776 addressing in accessible research. The Lancet. 383 (9913): 257-266.
777
778 Dellinger M, Carvan MJ, Klingler RH, McGraw JE, Ehlinger T. 2014. An
779 exploratory analysis of stream teratogenicity and human health using zebrafish
780 whole-sediment toxicity test. Challenges. 5(1):75-97.
781
782 Durda JL, Preziosi DV. 2000. Data quality evaluation of toxicological studies used
783 to derive ecotoxicological benchmarks. Hum Ecol Risk Assess. 6: 747-765.
784
785 The Economist. 2013. October. Trouble at the lab. Scientists like to think science
786 is self-correcting. To an alarming degree, it is not. Available from:
787 www.economist.com/news/briefing/21588057-scientists-think-science-self-
788 correcting-alarming-degree-it-not-trouble Downloaded 11/5/2016.
789
790 European Chemicals Agency. 2012. "How to Report Robust Study Summaries.
791 Practical Guide 3. Version 2.0."
792
793 EFSA supporting publication 2013:EN-511
794
795 Forbes V, Hall T, Suter GW. 2016. The Challenge: Bias is creeping into the
796 science behind risk assessments and undermining its use and credibility. Environ
797 Toxicol Chem. 35(5), 1068-1068.
798
799 Freedman LP, Cockburn IM, Simcoe TS. 2015. The economics of reproducibility
800 in preclinical research. PLoS Biol. 13, e1002165
801
802 Glasziou P, Altman DG, Bossuyt P, Boutron I, Clarke M, Julious S, Michie S,
803 Moher D, Wager E. 2014. Reducing waste from incomplete or unusuable reports
804 of biomedical research. The Lancet. 383 (9913): 267-276
805
806 Green JW, Springer TA Staveley JP. 2012. The Drive to Ban the NOEC/LOEC in
807 Favor of ECx is Misguided and Misinformed. Integrat Environ Assess Manag.
808 9:12-16.

Harris CA, Sumpter JP. 2015. Could the Quality of Published Ecotoxicological Research Be Better? Env Sci Technol. 49(16):9495-9496.

Harris CA, Scott AP, Johnson AC, Panter GH, Sheahan D, Roberts M, Sumpter, JP. 2014. Principles of sound ecotoxicology. Env Sci Technol. 48(6): 3100-3111.

Klimisch HJ, Andreae M, Tillmann U. 1997. A systematic approach for evaluating the quality of experimental toxicological and ecotoxicological data. Regul Toxicol Pharmacol. 25: 1-5.

Kolata G. 2013. Scientific articles accepted (personal checks, too). *New York Times*, April 7th.

Länge R, Hutchinson TH, Croudace C, Siegmund F, Schweinfurth H, Hampe P, Panter GH, Sumpter JP. 2001. Effects of the synthetic estrogen 17α-ethinylestradiol on the life-cycle of the fathead minnow (*Pimephales promelas*). Env Toxicol Chem. 20:1216-1227.

Leung J, Witt JDS, Norwood W, Dixon DG. 2016, Implications of Cu and Ni toxicity in two members of the *Hyalella azteca* cryptic species complex: Mortality, growth, and bioaccumulation parameters. Env Toxicol Chem. doi:10.1002/etc.3457

Mahoney MJ, 1977. Publication prejudices: An experimental study of confirmatory bias in the peer review system. Cognitive Therapy Res. 1(2):161-175.

Meyer JN, Francisco AB. A call for fuller reporting of toxicity test data. Integr Environ Assess Manag. 2013 Apr 1;9(2):347–8.

Moermond CTA, Kase R, Korkaric M, Ågerstrand M. 2016. CRED: Criteria for reporting and evaluating ecotoxicity data. Environ Toxicol Chem, 35:297–1309.

OECD 2012. Fish Toxicity Testing Framework. OECD Series on Testing and Assessment no. 171. ENV/JM/MONO (2012)16.

OECD 2006. Current approaches in the statistical analysis of ecotoxicity data: A guidance to application. Organisation for Economic Co-operation and Development. Report nr EVV/JM/MONO (2006) 18 Number 54. 1-147.

Ricaurte GA. 2003. Retraction. Science 301(5639): 1479b–1479.

Schneider K, Schwarz M, Burkholder I, Kopp-Schneider A, Edler L, Kinsner-Ovaskainen A, Hartung T, Hoffmann S. 2009. “ToxRTool”, a new tool to assess the reliability of toxicological data. Toxicol Lett. 189: 138-144.

Tufte ER. 1997. Visual Explanations: Images and Quantities, Evidence and Narrative. Cheshire, Conn., USA: Graphics Press. 156 p.

Tufte ER. 2001. The Visual Display of Quantitative Information. Cheshire, Conn. USA: Graphics Press. 197 p.

US EPA 2013. Risks of Cyfluthrin and *Beta*-cyfluthrin Use. Environmental Fate and Effects Division, Office of Pesticide Programs, United States Environmental Protection Agency. Washington, DC.

US EPA 1998. Reregistration Eligibility Decision (RED) for Dicofol.

US EPA. 2011. "Evaluation Guidelines for Ecological Toxicity Data in the Open Literature. Available Here: www.epa.gov/pesticides/science/efed/policy_guidance/team_authors/endangered_species_reregistration_workgroup/esa_evaluation_open_literature.htm."

van Dam RA, Harford AJ, Warne MStJ. 2012. Time to get off the fence: The need for definitive international guidance on statistical analysis of ecotoxicity data. Integrat Environ Assess Manag 8(2):242-245.

Warne MStJ, Batley GE, van Dam RA, Chapman JC, Fox DR, Hickey CW and Stauber JL. 2015. Revised Method for Deriving Australian and New Zealand Water Quality Guideline Values for Toxicants. Department of Science, Information Technology, Innovation and the Arts, Brisbane, Queensland. 50 pp. Available from: https://www.researchgate.net/profile/Michael_Warne/publication/287533532_Revised_Method_for_Deriving_Australian_and_New_Zealand_Water_Quality_Guideline_Values_for_Toxicants/links/56777c9508ae0ad265c5bd3d.pdf. Downloaded: 11/5/2016.

Weissgerber TL, Milic NM, Winham SJ, Garovic VD. 2015. Beyond bar and line graphs: time for a new data presentation paradigm. PLOS Biology 13: e1002128. DOI: 10.1371/journal.pbio.1002128.

FIGURES

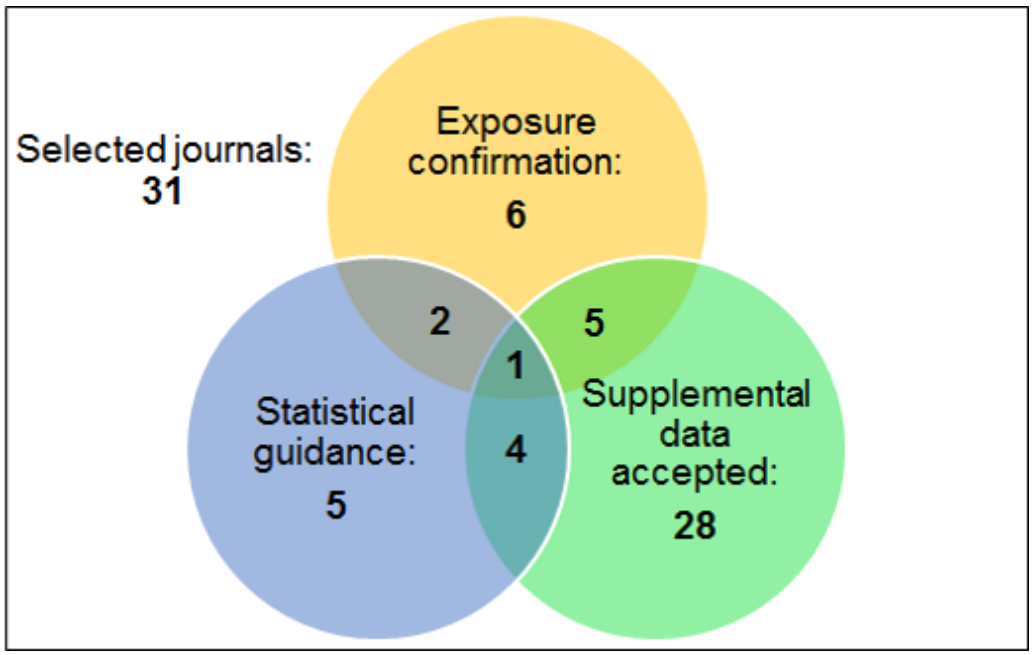

Figure 1: Number of journals that provide reporting requirements of 1) expectations around statistical analysis (blue circle), 2) confirmation of exposure concentrations (yellow circle), and 3) availability of supplemental data (green circle). Journals were selected using an ISI Web of Science search using the topic keyword 'ecotoxicology', and years published '2014 – 2015' (*n* = 172). Journals were further required to publish more than two 'ecotoxicology' articles in 2014 – 2015 (n = 31).

Table 1:

The model checklist provided below will assist authors and peer reviewers of ecotoxicology studies to improve their reporting and assessment.

Reporting Requirement	**Met?**
1. Test Compound Source and Properties	
Source and purity provided?	
Technical name?	
2. Experimental Design	
Hypotheses, if any, stated?	
Number of treatments and their exposure levels?	
Number and type of controls?	
Duration of exposures?	
Number of replicates?	
3. Test Organism Characteristics	
Name, source, and strain of species reported?	
Control performance criteria met?	
Husbandry protocols listed?	
4. Experimental Conditions	
General test conditions reported?	
Source and condition of media?	
Acclimation and feeding?	
5. Exposure Confirmation	
Clear statement of which samples were analyzed?	
Method LOD and LOQ provided?	
Nominal or measured used in subsequent analyses?	
6. Endpoints	
All endpoints monitored, regardless of response, provided?	
Clear definitions and measurement units provided?	
7. Presentation of Results and Data	
All data, regardless of statistical significance is discussed?	
Untransformed data provided?	
8. Statistical Analysis	
Statistical flowchart?	
Transformations justified?	
All outliers are reported	
Justification for model selection and variables?	
NOE-LOEC: power of test and percent change reported?	
ECx: Model estimates and confidence intervals provided?	
9. Raw Data	
Nominal and measured concentrations provided?	
Untransformed response by replicate available in some form?	