Original Article **Integrated Environmental Assessment and Management**
DOI 10.1002/ieam.1841

Future Needs and Recommendations in the Development of Species Sensitivity Distributions: Estimating Toxicity Thresholds for Aquatic Ecological Communities and Assessing Impacts of Chemical Exposures**

Scott Belanger†*, Mace Barron‡, Peter Craig§, Scott Dyer†, Malyka Galay-Burgos||, Mick Hamer#, Stuart Marshall††, Leo Posthuma‡‡, Sandy Raimondo‡, and Paul Whitehouse§§

† Procter & Gamble, Mason Business Center, P.O. Box 1093, Mason, OH 45040
‡ U.S. Environmental Protection Agency, Gulf Ecology Division, 1 Sabine Island Drive, Gulf Breeze, FL USA
§ Durham University Department of Mathematical Sciences, South Road, Durham DH1 3LE, UK
|| ECETOC, Avenue E. Van Nieuwenhuyse, 2 B-1160, Brussels, Belgium
Syngenta, Jealott's Hill Research Station, Bracknell RG42 6EY, UK
†† Safety & Environmental Assurance Centre, Unilever Colworth, Sharnbrook, Bedford, MK44 1LQ, UK
‡‡ National Institute for Public Health and the Environment, Centre for Sustainability, Environment and Health, P.O.Box 1, 3720 BA Bilthoven, the Netherlands and Radboud University Nijmegen, Department of Environmental Science, Nijmegen, the Netherlands
§§ Environment Agency, Red Kite House, Howbery Park, Wallingford, Oxon OX10 8BD, UK

Running Head: Advancing Species Sensitivity Distributions in ecotoxicology

Key Words: risk assessment, probabilistic, research needs, impact

*To whom correspondence may be addressed (Scott E. Belanger, Belanger.se@pg.com)

Submitted 18 February 2016; Returned for Revision 14 June 2016; Accepted 5 August 2016

ABSTRACT

A species sensitivity distribution (SSD) is a probability model of the variation of species sensitivities to a stressor, in particular chemical exposure. The SSD approach has been used as a decision support tool in environmental protection and management since the 1980s, and the ecotoxicological, statistical and regulatory basis and applications continue to evolve. This article summarizes the findings of a 2014 workshop held by ECETOC (the European Center for Toxicology and Ecotoxicology of Chemicals) and the UK Environment Agency in Amsterdam, the Netherlands on the ecological relevance, statistical basis, and regulatory applications of SSDs. An array of research recommendations categorized under the topical areas of Use of SSDs, Ecological Considerations, Guideline Considerations, Method Development and Validation, Toxicity Data, Mechanistic Understanding and Uncertainty were identified and prioritized. A rationale for the most critical research needs identified in the workshop is provided. The workshop reviewed the technical basis and historical development and application of SSDs, described approaches to estimating generic and scenario specific SSD-based thresholds, evaluated utility and application of SSDs as diagnostic tools, and presented new statistical approaches to formulate SSDs. Collectively, these address many of the research needs to expand and improve their application. The highest priority work, from a pragmatic regulatory point of view, is to develop a guidance of best practices that could act as a basis for global harmonization and discussions regarding the SSD methodology and tools.

INTRODUCTION

Chemicals are an integral element of human society and their production, use, and potential emissions are expected to grow in the future (UNEP 2013). This implies that continued attention to the safety and evaluation of chemicals is warranted for environmental protection (e.g., environmental standards, risk assessments), management (e.g., deciding what actions are required), and remediation (e.g., deciding what level of intervention or clean-up is acceptable or needed). A critical step in the assessment and control of chemicals in the environment is to understand their hazards and to estimate tolerable thresholds of risk. Various models and approaches are available to estimate chemical hazard levels, including Species Sensitivity Distribution (SSD) modeling. An SSD is a probability model of the variation of species sensitivities to chemical exposure. SSDs are increasingly used in ecological risk assessment and the derivation of environmental quality standards because they can be used to develop community-level thresholds, and have advantages over deterministic assessments that rely solely on application (uncertainty) factors applied to the most sensitive individual toxicity data point (OECD 1992; Wheeler et al. 2002; ECETOC 2014 wherein Posthuma provide a review). Some of the advantages of SSDs over application factors include:

- SSDs make full use of the knowledge on the toxicity of a substance;
- SSDs are explicit in expressing uncertainty;
- The shape and form of the SSD can inform the assessor about the behavior of the substance (e.g., steep slopes are often associated with specific modes of action);
- SSDs are probabilistic and as such are aligned with the paradigm of risk assessment as a probabilistic science (versus deterministic PNECs); and,
- The extrapolation process is flexible in that the level of protection can be defined relative to the percent of species potentially affected.

Management of chemicals in the environment usually includes comparison of expected exposures to a critical effect limit such as a Predicted No Effect Concentration for ecosystems (PNEC) (ECHA 2008). Concentrations below the PNEC are considered to have a negligible potential effect on the structure or function of an exposed ecosystem. When sufficient data are available a PNEC may be estimated as a low percentile of an SSD (Van Straalen and Denneman 1989). PNECs are most commonly deterministic and estimated by applying an application factor (AF) to the data derived from the most sensitive species tested (the actual AF being a function of the type of data, acute or chronic, and the number of species tested). When PNECs are estimated using SSDs, the extrapolation of laboratory test results to protect field populations and communities usually employs lower AFs (generally 1 to 5), while being somewhat flexible to account for the biological diversity present in and the statistical qualities of the SSD being considered (ECHA 2008). In this way, pragmatic implementation of SSDs as a regulatory tool often combines the probabilistic toxicity result (usually a small percentile from the toxicity distribution) with an AF (a deterministic approach). If the toxicity data set is sufficiently robust (e.g., it is built from tests on a large number of diverse species), the regulatory use may simply be the probabilistic toxicity result. In either case, these are based on environmental policy considerations for the regulatory jurisdiction doing the assessment.

Species Sensitivity Distributions have an established role in the assessment and management of risks posed by chemicals, and major developments around the world have provided relevant novel insights into their development and application. The formal adoption of SSDs for the derivation of environmental thresholds dates back to scientific- and policy milestones of 1985 in the United States and 1989 in Europe (Stephan et al. 1985; Van Straalen and Denneman 1989). In 2001, SSDs were evaluated intensively for the derivation of European environmental quality standards (EC 2001). In 2002 a comprehensive overview of the principles and practices of SSD use on an international basis was made (Posthuma et al. 2002), followed by a recent updating review on the use of SSDs with

particular focus on environmental quality criteria by Del Signore et al. (in press). The latter review confirms and expands on our analyses, further underpinning the conclusions as well as the needs recommendations for future developments and use of SSDs.

Here we summarize the major findings of a workshop sponsored by the European Centre for Ecotoxicology and Toxicology of Chemicals (ECETOC) and the UK Environment Agency held in February 2014 in Amsterdam, The Netherlands (ECETOC 2014). Forty experts from academia, business, and government reviewed the state of the science for estimating toxicity thresholds for aquatic ecological communities using SSD modeling, and considered advances in statistical, ecotoxicological, and ecological science applicable to SSDs that have occurred since a similar workshop was held in London in 2001 (EC 2001). New approaches or refinements to current applications of SSD modeling were evaluated against current methods in which SSDs are used in the context of environmental protection and management. The aim of this paper is to provide an overview of the findings, conclusions and recommendations of the workshop, in order to derive future recommendations given the state of the science and the needs for decision support.

DERIVATION OF SPECIES SENSITIVITY DISTRIBUTIONS

Predictive risk and retrospective impact assessment of chemicals requires estimation of the toxicity thresholds of chemicals for aquatic communities as an integral aspect of defining environmental hazard. Of the available tools used in hazard and risk assessment, SSDs provide a particularly informative approach because they explicitly relate the intensity of chemical pressure (e.g., the concentration) to ecological impacts (the proportion of species at risk). Currently, hazard is most frequently predicted using concentration–effect data from single species laboratory toxicity tests that measure effects on individuals and populations. Typically, responses of individuals include survival (applicable to acute and chronic testing), growth and reproduction endpoints for invertebrates, fish,

amphibians and macrophytes. Population responses such as growth rate for microinvertebrates, bacteria, algal and cyanobacteria tests are used in acute and chronic exposures. However, protection goals are generally broader than those covered by endpoints derived from laboratory toxicity testing and focus on populations, communities, and ecosystems. There is growing interest in moving from hazard levels derived from individual toxicity tests to the use of SSDs, which can better be used to estimate potential hazards to communities. Note that while SSDs include multiple species, they are the compilation of individual species responses and typically do not include inter-specific interactions (predation, competition) or ecosystems processes (nutrient cycling, energy flow).

The statistical methods and underlying scientific foundation supporting the use of SSD models and the versatile use of these in environmental protection, assessment and management were reviewed by Posthuma as discussed at the Workshop and reported earlier (ECETOC 2014). Briefly, the SSD method assembles single species toxicity data to predict a hazardous concentration (HC*p*) affecting a certain percentage (*p*) of all the species in a distribution, or to estimate the toxic pressure, expressed as the potentially affected fraction (PAF) of species, exerted on an assemblage from an observed or expected exposure concentration. SSDs can be constructed using either acute or chronic test data, depending on data availability, and they can be related to the protection goal. In comparisons amongst chemicals, SSDs derived from ecotoxicity data can have different positions (intercept) and shapes (slope) which has implications for the HC*p* of a chemical and the toxic pressure of an environmental sample. The higher the HC*p* of a chemical, the lower is its ecotoxic potential to induce impacts. Greater toxic pressure is indicated by a larger PAF for a contaminated sample. The potential for expected impacts for tested species and impacts on aquatic communities is therefore assumed to be greater when toxic pressure increases.

SSDs are constructed with the aim of predicting acute or chronic toxicity, although these are usually dealt with separately. Single species data for acute toxicity (expressed as median lethal or

effective concentrations [LC50, EC50]), or estimates for chronic effects (expressed as, no-observed-effect concentrations [NOECs], Chronic Values [defined as the geometric mean of the NOEC and LOEC, Lowest Observed Effect Concentration], and EC10s) for several species are fitted to one or more cumulative distribution functions followed by evaluation and choice of the best model. The cumulative distribution function is often assumed to be lognormal or log-logistic (Awkerman et al. 2013; Posthuma et al. 2002). Other distributions have been used and can also have utility (ANZECC and ARMCANZ 2000; Warne et al. 2015). A typical approach uses the 5th percentile of the distribution of acute or chronic effects to derive toxicity thresholds or environmental quality criteria that should ensure that the specified level of protection is achieved. The estimation of the toxic pressure (PAF of species) given an ambient exposure allows the use of any endpoint (e.g., NOEC, EC10, EC50, LC50), depending on the expected level and duration of exposure. Similarly, the estimated toxic pressure can yield assessment outcomes such as a PAF_{NOEC}, or a PAF_{EC50} that specify the fraction of species exposed above their NOEC or EC50, respectively. For example, an ambient exposure might predict that 50% of the species are exposed above their NOEC while at the same time 20% of species are exposed above their EC50.

One of the principle advantages of probabilistic SSDs over deterministic application factors is the opportunity to express uncertainty in the point estimate (HC*p*) as additional information for the risk assessor to judge the utility of the estimated threshold. Typically, the HC5 will be accompanied by a confidence limit that conveys knowledge of the shape of the statistical distribution of toxicity values and their variance. By addressing critical data that appear to strongly influence the shape of the distribution (often at the tails of tolerance and sensitivity) the risk assessor can understand the impact of particular data on the HC*p* and the confidence interval around it.

ECOLOGICAL, STATISTICAL, AND REGULATORY CONSIDERATIONS

Since the sensitivity of all the species that might be exposed to a chemical cannot be known, extrapolation needs to be done from the data available. ECETOC (2014) discussed that scientifically sound extrapolation approaches based on SSDs to derive toxicity threshold concentrations should provide a more useful and transparent assessment of risks than a deterministic approach using generic factors applied to single species aquatic toxicity test data. The SSD methodology is a valuable regulatory and management tool since it can provide greater insight into the potential effects of a particular level of exposure compared to the deterministic application factor method, enabling better problem definition and decision support.

Regulatory tools such as SSD modelling are useful if they strike a balance between being overly cautious and under-protective. Being overly protective can lead to unnecessary mitigation costs and stifle innovation whereas under protection may result in environmental degradation (ECETOC 2014). A prospective risk assessment conducted in the context of environmental protection needs to establish that there will be acceptable risk at the criterion concentration (e.g., Predicted No Effect Concentration for Ecosystems [PNEC], Environmental Quality Standard [EQS], or Regulatory Acceptable Concentration [RAC]). In contrast, retrospective impact assessment uses diagnostic tools to identify the cause of existing adverse effects, using SSDs to quantify expected chemical impacts compared to other stressors (De Zwart et al. 2006). When sufficiently large datasets are available, the risk of errors is reduced, while uncertainty on expected protection or impact prediction declines. In such cases, SSD modelling provides a mechanism for quantifying the relationship between chemical pressure and impact that takes account of uncertainty due to differences in sensitivity between species. When datasets are small, uncertainty is greater and consequently the more cautious deterministic approach may be more appropriate. That is, the criterion is derived from the available data combined with an application factor. Under conditions of small data sets (e.g., few species tested) or lower data quality, a higher application factor is implied and appropriate for the deterministic assessment. Similarly, the

size of an assessment factor applied to an SSD will vary (minimum of 1) according to the uncertainty in the hazard estimation.

Requirements for consideration of an SSD approach vary across regulatory jurisdictions (e.g., by national regulatory authority), regulatory frameworks for specific compound classes (e.g., pesticides covered under US FIFRA or EU PPP D [1107/2009]) or intended use in an assessment framework (e.g., water quality standards or chemical-specific risk assessments). Table 1 provides an overview of representative (not exhaustive) considerations in several frameworks. It is interesting to note the variation in species coverage, treatment of multiple data on the same species used as SSD input, and application of statistical principles that are applied. The most recent guidances on SSD use for assessing hazards of chemicals (ECHA 2008) and plant protection products (EFSA 2013) are not surprisingly the most complete across all the facets to be considered. These guidances are consistent with discussions in Europe in the previous decade (EC 2001; ECETOC 2008) and form the basis of subsequent national and international guidance used in setting water quality criteria as well (e.g., CCME 2007; EC 2011).

ECETOC (2014) cautioned that continued validation of predictions made using SSDs against a reference tier, such as field and mesocosm data, is required to ensure that a threshold derived from an HC*p* (sometimes coupled with an application or safety factor) or a PNEC (Predicted No Effect Concentration) has ecological relevance (see also Versteeg et al. 1999; Posthuma et al. 2002). A new development is the advent of the SSD approach applied to field data rather than field data being regarded as a separate line of evidence (Kwok et al. 2008). The results of any extrapolation process (including SSDs) should always be critically assessed based on all available knowledge on the substance and related substances, such as their mode of action and other lines of evidence including field and mesocosm data. Use of the SSD methodology should yield more generally conservative estimations of hazard (i.e., lower predicted effect concentrations) and thus more readily acceptable

results in most regulatory contexts than those obtained from mesocosm-based methods (Versteeg et al. 1999). Differences remain across regulatory jurisdictions on this aspect (for example, Canadian and Australian regulatory decisions would place increased emphasis on mesocosm results if conducted following sound statistical, biological and ecological principles; ANZECC 2000; CCME 2007). Mesocosms and field studies will remain valuable tools for evaluating the accuracy of SSD predictions because of the inherent interactions among populations and communities that are not inherent in single species tests. Further, as acknowledged in many other venues, mesocosms often have the additional advantage of utilizing more realistic field exposures (Giddings et al. 2002).

A new development in the use of SSDs is an emerging interest in using field data based on population abundance and biomass as alternatives to toxicity estimates in the laboratory (Leung et al. 2005). Field-based SSDs may allow an expansion of taxonomic coverage and thus provide insight into responses for taxa less easily tested in the laboratory but that exist temporally in the same space. On the other hand, intra- and inter-specific interactions as well as multiple-stress responses are certainly involved in field assessments. Therefore, the interpretation or meaning of the SSD may change compared with assessments based solely on laboratory single species toxicity tests.

Multiple statistical approaches are available for SSD modeling and high uncertainty can arise in cases of limited taxa diversity (ECETOC 2014). To address data gaps in taxa diversity, the hierarchical SSD (hSSD) was developed as a novel approach and discussed by Craig and colleagues (Craig et al. 2012; Craig 2013; ECETOC 2014). This can be used to predict thresholds for defined species assemblages using knowledge of the general trends in how species sensitivity is related to their taxonomic distance. Other methods for addressing data gaps in taxa diversity include the U.S. EPA Web-ICE tool (www.epa.gov/ceampubl/fchain/webice/; Raimondo et al. 2016) which uses interspecies correlation estimation models to estimate toxicity for taxa with limited data (Awkerman et al. 2013). The U.S. EPA Web-ICE tool also explored interspecies toxicity estimation as a function of taxonomic

distance and showed the phenomenon is generally important. While the investigations do not aim to assess the influence of chemical class on the relationship, the fact that many modes of action are present in the database suggest it is a generalized phenomenon. Traditional statistical approaches, Web-ICE, and the hSSD prototype were compared and contrasted in ECETOC (2014) using case studies involving the surfactant linear alkylbenzene sulfonate and the insecticide chlorpyrifos. Three distinct regulatory applications associated with the use of SSDs are evident:

1. The derivation of generic protective threshold concentrations applied to many different locations, perhaps over very large geographical regions. These are assumed to offer sufficient protection everywhere, even in the most sensitive systems.

2. The derivation of scenario-specific protective thresholds that more closely reflect local conditions (e.g., constrained to resident species or for a certain water quality condition), but which may not be transferable from one place to another.

3. Identifying the causes of biological impact ('diagnosis') or expected impact magnitudes of existing or expected (mixture) contamination, in order to inform the need and focus for any remedial or management action.

The first 2 applications are protective and thus will tend to include a certain amount of precaution, while in contrast the third needs to be predictive. The diagnostic use of SSDs has recently been summarized by Posthuma et al. (2016).

RESEARCH NEEDS

The overview of SSD practices as discussed during the workshop has shown that SSDs currently have a significant influence on national and international decision making regarding assessments of chemical exposure to ecosystems. It is evident from review of current applications of

SSDs in regulatory decision-making that better understanding of the state of the science and answers to frequently asked questions would encourage best practices in the use of SSDs by regulators, risk assessors, and risk managers. Although expert judgement has a role in the interpretation of SSD models, a compilation of current best practices would provide a valuable compendium of regulatory experiences beneficial to countries seeking to derive their own environmental quality standards or to scientists seeking to understand the significance of emerging chemicals or new applications of existing chemicals on ecosystems. An array of modelling tools has extended the statistical evaluation of SSD "quality" that builds upon progressively better and more available input data as a result of global chemical management programs (e.g., OECD HPV [High Production Volume] Challenge program, European REACH, Canadian Categorization of the Domestic Substances List and others). According to ECETOC (2014) the use of species sensitivity distributions in ecological diagnostics links policy targets on ecological integrity, monitoring data, SSD modeling and landscape-level mixture impact diagnosis. Therefore, research that builds a stronger scientific foundation is preferable to work focused narrowly on a single species or taxa.

Specific research needs were identified in the workshop that would augment the application of SSDs in most circumstances: The research needs were divided into the following themes: use of the SSD, ecological considerations, guideline considerations, model development and validation, toxicity data, mechanistic approaches, and uncertainty (Table 2). The most important of these are highlighted here.

1. Tools for regulatory decision making should be given high priority with particular focus on i) SSDs for chronic toxicity, ii) validating HC5s with mesocosms and real ecosystems, and iii) maximising the use of available data, e.g. by applying weighting criteria.

   Rationale: the most potentially influential use of SSDs is establishing safe concentrations for ecosystems associated with long term, low level exposure to chemicals, therefore assessments

based on chronic exposures are essential. However, the use of SSDs in general should be somewhat more conservative (i.e., predict lower hazardous concentrations) for routine use than higher tier studies (e.g., mesocosms). Higher tier studies should still behave consistently with predictions provided by SSDs (Versteeg et al. 1999). Roles for acute SSDs can also be relevant, and in some situations critical, such as short term pesticide exposures.

2. Mechanisms to maximize the use of available data should be further developed, e.g. by applying weighting criteria to broaden taxonomic coverage and use of non-GLP (Good Laboratory Practice) studies.

Rationale: The majority of standardized toxicity tests focus on relatively few species. Taxonomic coverage is a key facet of developing SSDs and non-standard tests are increasingly used as input. These are also most often not performed under a GLP framework. Weighting or valuing different types of studies should be explored to maximize the use of all high quality data that are available. The inclusion or exclusion of studies has been shown to be one of the single largest contributors to variance in PNEC and SSD derivation (Hahn et al. 2014). A recent Pellston Workshop entitled "Use of Ecotoxicology in Regulatory Decision-Making" in Shepherdstown, WV USA from August 30-September 4, 2015) was held to frame this issue and propose solutions (Hanson et al. In Press).

3. Further development of tools for assessing mixtures of chemicals.

Rationale: Aquatic and sediment environmental exposures are rarely to single chemical or stressor insults and are more commonly to mixtures. Methods to perform aggregate and cumulative assessments are needed for the future as mixture assessments are increasingly demanded by the stakeholders. Effluent toxicity assessments address this to a degree but SSD-

based mixture assessments are possible if mode of action and theories of concentration addition and independent action can be accounted for (Kapo et al. 2014).

4. Trait-based SSDs appear to offer advantages over conventional taxonomic based approaches, but there is currently no practical application.

   Rationale: This continues to be a developing science in ecotoxicology. It is likely that responses to chemicals are in part based on ecological traits (much like their classifications in feeding or trophic ecology) with some trait types more sensitive to certain types of exposures than others (Pilière et al. 2014). In this context, traits are morphological, physiological, or phenotypic heritable features that are measurable at the level of the individual. Trait-based ecotoxicology proposes to link stress response patterns of species to effects at the level of ecosystem properties. Trait-based SSDs then would focus on groupings, other than species, as inputs to the SSD thereby acknowledging the importance of preserving organism functional roles in addition to classical biodiversity.

5. SSDs for more taxa including plants and, possibly, micro-organisms.

   Rationale: It is well established that photosynthetic micro-algae are frequently more sensitive than fish or invertebrates (Jeram et al. 2005) but are sometimes not considered in SSD formulation. Photosynthetic and non-photosynthetic microbes, aquatic macrophytes and plants play crucial roles in ecosystem structure and function, therefore, including these species in SSDs more frequently may improve robustness of predictions.

6. Development of a more scientifically critical role for cheminformatic approaches.

   Rationale: Future environmental toxicology approaches should be able to take advantage of the large efforts on-going in programs such as the US NRC "Toxicity Testing in the 21st Century"

(NRC 2007). Cheminformatics is the strategic use of computer and informational techniques applied to a range of problems in the field of chemistry including those of drug discovery, development of *in silico* models, and relating key chemical attributes to the potential for hazard. Environmental scientists generally have a strong appreciation for physical-chemical attributes in testing and assessment that will bridge well to cheminformatics. How SSDs approaches can take advantage of cheminformatics developments should be explored.

7. Focus on sensitive groups.

Rationale: A better understanding of the frequency of bi-modality in SSDs is needed (i.e., when one taxonomic group is more sensitive compared to others) and how to further incorporate this into assessment methodologies is needed. Certain groups of chemicals may even benefit from a greater focus on sensitive subgroups, for example micro-algae to anti-microbials, as a stronger basis for extrapolation for environmental protection.

8. The usefulness/applicability of SSDs for defined communities.

Rationale: Approaches of the h-SSD form provide some unique advantages to probe relationships between available studies used as SSD inputs and actual distributions of species based on taxonomy observed in the field (Craig et al. 2012; Craig 2013).

9. Internal dose (CBB or critical body burden)-based approaches have potential to incorporate mechanistic toxicokinetic/toxicodynamic modelling evidence that could help explain sensitivity differences between taxa/traits.

Rationale: Critical body burden concepts allow a technically defensible determination of exposure to chemicals at the target organ of interest resulting in acutely or chronically toxic effects (McCarty et al. 1992; McElroy et al. 2011). CBB approaches have generally been

investigated for organic compounds and are not only more mechanistically-based, a laudable goal in any toxicological investigation, but also have the attractive feature of providing insight into mixture assessments. Greater emphasis on developing CBB for algae and invertebrates would need to be undertaken as fish have been the primary group of interest until now. Development of CBB data for a broader array of taxa would need to be addressed in order for body burdens to be utilized in SSD development. SSDs based on CBB could conceivably provide deeper insight into true hazardous concentrations versus those based on external concentrations only. A new framework for regulatory use (e.g., SSDs per taxonomic group) may be needed as modes of action for single chemicals may not translate well across taxa

10. Quantifying uncertainty as an alternative to standard application factors.

Rationale: It is acknowledged that this will be a challenge for any regulatory framework, however, it is consistent with the goals of risk assessment which is fundamentally probabilistic in nature. Research is needed to ascertain the relationship of statistical uncertainty with deterministic application factors typically applied to small data sets. Improvements to the role of application factors, even as they are applied to SSD results, due to variation in SSD quality, are also warranted.

11. What level of confidence do current SSD criteria continue to provide

Rationale: Through the development of more unified global best practices, the means to value the varying levels of quality resulting from SSD methods may become clear. Treatment of data (multiple studies on the same species, different endpoints utilized even for the same species), taxonomic coverage (breadth of species, species choices), statistical models used, and how these affect HC5 predictions and their uncertainties is essential for long term support of the tool.

ACKNOWLEDGEMENTS

This article summarizes the primary outcomes from a workshop entitled "*Estimating toxicity thresholds for aquatic ecological communities from sensitivity distributions*", held 11-13 February 2014, in Amsterdam, the Netherlands. The objectives of the workshop were to: (1) study and where possible improve the ecological relevance of SSDs, (2) collate, compare and where possible improve statistical approaches for SSD modeling, and (3) describe and evaluate regulatory applications of SSDs. The Workshop was supported by the European Centre for Ecotoxicology and Toxicology of Chemicals (ECETOC), and the Environment Agency for England and Wales. LP was funded by the CEO of RIVM under project nr. S/015031. This paper has been reviewed according to U.S.EPA guidelines, and the opinions expressed in this work are those of the authors and do not represent the policies or opinions of the U.S. EPA.

DATA AVAILABILITY

No actual data are presented in this paper.

REFERENCES

ANZECC (Australian and New Zealand Environment and Conservation Council) and ARMCANZ (Agriculture and Resource Management Council of Australia and New Zealand). 2000. Australian and New Zealand Guidelines for Fresh and Marine Water Quality Volume 2 Aquatic Ecosystems — Rationale and Background Information (Chapter 8). 678 pp.

Awkerman JS, Raimondo S, Jackson CR, Barron MG. 2013. Augmenting aquatic species sensitivity distributions with interspecies toxicity estimation models. *Environ Toxicol Chem* 33:688-695.

CCME (Canadian Council of Environmental Ministers). 2007. A Protocol for the Derivation of Water Quality Guidelines for the Protection of Aquatic Life (Draft March 14, 2007). Part 1. Canadian Water Quality Guideline Framework. Ottawa, Canada 36pp.

Craig PS, Hickey GL, Luttik R, Hart A. 2012. Species non-exchangeability in probabilistic ecotoxicological risk assessment. *J Royal Stat Soc. Series A: Statistics in Society* 175:243-262.

Craig PS. 2013. Exploring Novel Ways of Using Species Sensitivity Distributions to Establish PNECs for Industrial Chemicals: Final Report to Project Steering Group. Downloaded from: http://dro.dur.ac.uk/13383 (last accessed 10/02/2016)

De Zwart D, Dyer SD, Posthuma L, Hawkins CP. 2006. Predictive models attribute effects on fish assemblages to toxicity and habitat alteration. *Ecol App* 16:1295-1310.

Del Signore A, Hendriks AJ, Lenders HR, Leuven RS, Breure AM. Development and application of the SSD approach in scientific case studies for ecological risk assessment. *Environ Toxicol Chem* (in press).

EC (European Commission). 2001. EC Expert Consultation Workshop on Statistical Extrapolation Techniques for Environmental Effects Assessments. Report of the London Workshop, 17-18th January, 2001; European Commission, Joint Research Center, Institute for Health and Consumer Protection, European Chemicals Bureau, Ispra (VA), Italy. 9 p.

EC (European Commission). 2011. Technical Guidance for Deriving Environmental Quality Standards. CIS Guidance Document No. 27, Technical Report – 2011 – 055.

ECETOC (European Center for Toxicology and Ecotoxicology of Chemicals). 2008. Workshop on the probabilistic approaches for marine hazard assessment, 18-19 June 2008, Oslo, Norway. Brussels, Belgium. Workshop Report No. 15, 83 pp.

ECETOC (European Center for Toxicology and Ecotoxicology of Chemicals). 2014. Estimating toxicity thresholds for aquatic ecological communities from sensitivity distributions. 11-13 February 2014, Amsterdam. Workshop Report No. 28. European Centre for Ecotoxicology and Toxicology of Chemicals. Brussels, Belgium.

ECHA (European Chemicals Agency). 2008. Guidance on information requirements and chemical safety assessment Chapter R.10: Characterization of dose [exposure]-response for environment. May, 2008. 65 p. Downloaded from: http://echa.europa.eu/reach_en.asp.

EFSA (European Food Safety Authority). 2013. Scientific Opinion, Guidance on tiered risk assessment for plant protection products for aquatic organisms in edge-of-field surface waters, EFSA Panel on Plant Protection Products and their Residues (PPR). *EFSA J* 2013 11(7):3290, 268p.

Giddings JM, Brock TCM, Heger W, Heimbach F, Maund SJ, Norman SM, Ratte HTA, Schaefers C, Streloke M. 2002. Community-Level Aquatic System Studies – Interpretation Criteria. Society of Environmental Toxicology and Chemistry, SETAC Press. 60p.

Hahn T, Diamond J, Dobson S, Howe P, Kielhorn J, Koennecker G, Lee-Steere C, Mangelsdorf I, Schneider U, Sugaya Y, Taylor K, Dam RV, Stauber JL. 2014. Predicted no effect concentration

derivation as a significant source of variability in environmental hazard assessments of chemicals in aquatic systems: an international analysis. *Integr Environ Assess Manag* 10:30-36.

Hanson ML, Wolff BA, Green JW, Kivi M, Panter GH, St.J.Warne M, Ågerstrand M, Sumpter JP. 2016. How we can make ecotoxicology more valuable to environmental protection. *Sci Total Environ* (in press).

Jeram S, Sintes JM, Halder M, Fentanes JB, Sokull-Klüttgen B, Hutchinson TH. 2005. A strategy to reduce the use of fish in acute ecotoxicity testing of new chemical substances notified in the European Union. *Regul Toxicol Pharmacol* 42: 218-24.

Kapo KE, Holmes CM, Dyer SD, De Zwart D, Posthuma L. 2014. Developing a foundation for eco-epidemiological assessment of aquatic ecological status over large geographic regions utilizing existing data resources and models. *Environ Toxicol Chem* 33:1665-1677.

Kwok, KWH, Bjorgesæter A, Leung KMY, Lui GCS, Gray JS, Shin PKS, Lam PKS. 2008. Deriving site-specific sediment quality guidelines for Hong Kong marine environments using field-based species sensitivity distributions. *Environ Toxicol Chem* 27:226-234.

Leung KMY, Bjorgesater A, Gray JS, Li WK, Lui GCS, Wang Y, Lam PKS. 2005. Deriving sediment quality guidelines from field-based species sensitivity distributions. *Environ Sci Technol* 39:5148-5156.

McCarty, LS, Mackay D, Smith AD, Ozburn GW, Dixon DG. 1992. Residue-based interpretation of toxicity and bioconcentration QSARs from aquatic bioassays: Neutral narcotic organics. *Environ Toxicol Chem* 11:917-930.

McElroy AE, Barron MG, Beckvar N, Kane SB, Driscoll T, Meador JP, Parkerton TF, Preuss TG, Stevens JA. 2011. A review of the tissue residue approach for organic and organometallic compounds in aquatic organisms. *Integ Environ Assess Manag* 7:50-74.

NRC (National Research Council). 2007. Toxicity testing in the 21st century : a vision and a strategy. National Academies Press, Washington, DC.

OECD (Organization for Economic Cooperation and Development). 1992. Report of the OECD Workshop on Extrapolation of Laboratory Aquatic Toxicity Data to the Real Environment. OECD Monograph No. 59. Paris, France. 45p.

Pilière A, Schipper AM, Breure, AM, Posthuma L, De Zwart D, Dyer SD, Huijbregts MAJ. 2014. Comparing responses of freshwater fish and invertebrate community integrity along multiple environmental gradients. *Ecolog Indicators* 43:215-226.

Posthuma L, Suter GW, Traas TP. 2002. Species sensitivity distributions in ecotoxicology. Lewis Publishers, CRC Press. Boca Raton, Florida.

Posthuma L, Dyer SD, De Zwart D, Kapo K, Holmes CM, Burton GA. 2016. Eco-epidemiology of aquatic ecosystems: Separating chemicals from multiple stressors. *Sci Total Environ*. In press.

Stephan CE, Mount DI, Hansen DJ, Gentile JH, Chapman GA, Brungs WA. 1985. Guidelines for Deriving Numerical National Water Quality Criteria for the Protection of Aquatic Organisms and Their Uses. National Technical Information Service, Washington D.C. USA. 104 p.

UNEP (United Nations Environmental Programme). 2013. UNEP *Global chemicals outlook. Towards sound management of chemicals*; United Nations Environment Programme: Geneva, Switzerland. p. 248

Raimondo S, Lilavois CR, Barron MG. 2016. Web-based Interspecies Correlation Estimation (Web-ICE) for Acute Toxicity: User Manual Version 3.3. EPA/600/R-15/192, June 2016. 23p.

Van Straalen NM, Denneman CAJ. 1989. Ecotoxicological evaluation of soil quality criteria. *Ecotoxicol Environ Saf* 18:241-51.

Versteeg DJ, Belanger SE, Carr GJ. 1999. Understanding single-species and model ecosystem sensitivity: data based comparison. *Environ Toxicol Chem* 18:1329-1346.

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.

Wheeler JR, Grist EPM, Leung KMY, Morritt D, Crane M. 2002. Species sensitivity distributions: data and model choice. *Mar Poll Bull* 45:192-202.

Table 1. Examples of the use of Species Sensitivity Distributions in several regulatory frameworks for the purposes of chemical risk assessment and formulation of water quality criteria or standards. The table is non-exhaustive and other frameworks are used by hazard assessors globally.

		Standard for Acceptance in Regulatory Framework			
Facet of SSD Development	**Factors to be considered**	**ECHA 2008**	**USEPA Ambient Water Quality Criteria, Stephan et al. 1985**	**CCME 2013**	**ANZECC and ARMCANZ (2000) supplemented by Warne et al. (2015)**
Type of regulatory use		Chemical hazard assessment under REACH	Development of chemical discharge criteria and impaired water assessments	Development of chemical discharge criteria, also used in chemical hazard assessment	Development of chemical discharge criteria
Overall quality of the database	Information or data source	Klimisch scoring, preferably in IUCLID5, documented	Data available in typed and dated form (publication, manuscript, letter, memorandum, etc.); supporting information indicates acceptability results are probably reliable	Use of acceptable laboratory practices in design and execution of tests. Each study is classified as primary, secondary, or unacceptable, based on detailed inclusion criteria.	Scored using a reliability assessment system specific to Australia and New Zealand
	Are data generated from true chronic studies	Required	Acute data required; chronic data are needed for Acute:chronic ratio (ACR) determination	Required for primary data	Required for chronic SSD formulation; note that acute SSDs can be generated as well
	Chronic studies cover sensitive life stages	Required	Not required	Required for primary data	Not defined; notation of life stages is required
	Endpoints used as input are the lowest NOECs or ECx values of the endpoints measured in relevant studies	Required	EC50 and LC50 data required; NOEC/LOEC data are needed for ACR determinations	ECx prefered over NOECs (x 10 or less)	ECx is preferred over NOECs from all studies and endpoints on each species
	Treatment of endpoint data for multiple tests on same species	Not specified	Always uses survival or immobility endpoints thus records are similar by definition; LC50 and EC50	Median value for comparable records for the same endpoint when more than two data points are available; if only	Data for each species and endpoint expressed as a geometric mean and

			estimates calculated at the genus level as geometric means	two the geometric mean is used to represent the average species effects endpoint	the lowest value per species is used
	Use of Data on Most Sensitive Endpoints	NOEC conclusions from the quoted studies should represent the most sensitive endpoint for the test	See above	See above	See above
Taxonomic Groups Considered	Fish	At least two species	Salmonid is required; second species of commercial or recrational importance is recommended (a second fish is required)	Three species including one coldwater species (salmonid), one warmwater species, and one other.	No specific organism types are required; at least 4 phyla needed
	Additional vertebrates	Not required	Third chordate family required (amphibian or third family of fish)	Amphibian highly desired, but not required	Not specified
	Crustaceans	At least one species	Required	Required	Not specified
	Insect	At least one species	At least one species	Mayfly, stonefly or caddisfly preferred but not required	Not specified
	Additional invertebrates	At least one additional phylum not represented by Insects or Crustaceans	At least one more family in a phylum other than Arthropoda or Chordata and at least one more family in any order of insect or any phylum not already represented	Two additional invertebrates required	Not specified invertebrates
	Algae (number unspecified)	Required but number unspecified	Not used	One species of a plant or alga is required; three required if indications exist that photosynthetic organisms are sensitive	Not specified
	Higher plants (number unspecified)	Required but number unspecified	Not used	One species of a plant or alga is required; three required if indications exist that photosynthetic organisms are sensitive	Not specified
Minimal Sample Size	Total number of species in SSD	10 NOECs, preferably more than 15 for different species	8 different families required	At least 10, preferably 15 different species	Minimum of 5 species from at least 4 taxonomic groups, preferably more than 5 species
Statistical Fit to A Distribution	Use of Underlying Distribution	Confirm model choice, flexible for data, but	Log triangular required applied to the four most	Confirm model choice	Burr Type III distribution

		lognormal and log logistic identified as most common	sensitive genera		recommended (note this includes log-normal, log-logostic and log-triangular)
	Statistical Goodness of Fit	Confirm by appropriate GoF test	Not considered	Confirm GoF	Confirm GoF
	Conclusion	Provided overall statement as fit for purpose	Generally should discuss	Generally should discuss	Generally should discuss
Estimated Parameter	HC5 and Confidence interval	HC5 with 50% CI derived and provided	HC5 is derived as input into further calculations to establish the water quality criterion	HC5 with 95% CI derived and provided	HC1, HC5, HC10 and HC20 with 50% CI derived and provided to address various protection targets
NOEC values below the HC5	Discuss values that fall below the HC5	Required	If economically or recreationally important species fall below the HC5, the criterion will be lowered to protect those species.; although algae and plants are not included in the SSD, algae and plant toxicity data are compared to the HC5	Discuss	HC5 should be less than the chronic effect concentration for high value or keystone species
Distribution of trophic levels within the SSD	Discuss trophic level influences	Assess distribution of trophic level within the chosen distribution; use multiple curves if bi- or multi-modal	Not required	Assess distribution of trophic level within the chosen distribution; use multiple curves if bi- or multi-modal	Assess distribution of trophic level within the chosen distribution; use multiple curves if bi- or multi-modal
Knowledge of the Mode of Action	Discuss	Required	Indicated in documentation	Required	Indicated in documentation

Table 2. Major categories of work that could improve the long term application, usability, and interpretation by risk assessment practitoners.

Research area	**Description**
Uses of SSD	Collate and review the uses of SSDs for purposes other than estimating the HC5 (e.g. using the entire SSD for probabilistic risk assessment and deriving other values (say HC50) for trigger management action).
Ecology	Investigate whether an approach which allows better extrapolate to all ecosystems is viable.
	Compare trait-based SSDs with traditional strictly taxonomic-based SSDs, and to define what traits are most relevant to SSD generation. Alternative approaches should be explored, including focusing on sensitive taxa rather than broadly populating an SSD. However, there is uncertainty of what the sensitive taxa will be for many substances. A sensitive species approach may require novel methods development, including integrating chemical structure, genomic, traits and MOA information.
	Compare SSD-based approaches to the use of generic AF values under different scenarios of data richness, and the need to explore uncertainty in relaxed (10 species/8 taxa group) requirements versus AF uncertainty and conservatism. Determination of the ecology and composition of representative ecosystems should inform requirements for taxa composition in SSDs. SSD-based estimates determined from various approaches and data richness scenarios should be compared to field data, and field monitoring should be performed to verify SSD-based predictions of community level effects.
	(Further) Develop a model that takes account of the number and type of species in a community and that shows the consequences/reliability of the results. Establish what validity criteria are needed.
	Determine what additional ecological knowledge needs to be included to add value for the risk assessors.
Guidelines	Develop a formal and transparent decision tree approach that is inclusive of the available data, and that considers the generic or specific use of SSDs in environmental protection and management.
	Develop guidelines on how to deal with data quality (of the input data on species sensitivities, or sometimes functions sensitivities).
	Develop guidance on the use of non-standard test species.
	Develop guidance on which methods and tools can be used to generate SSDs – this requires sensitivity analysis, identification of causes of differences, etc.
Model development and validation	Investigate the limitations of the models and whether they are fit for the purpose for which they are used.
	Evaluate the viable methods for incorporating all relevant data in SSDs
	Further validation of SSDs derived from laboratory data against field and mesocosm studies is required, as is guidance on the different approaches (including their limitations) that can be taken.

	Further validation for extrapolations that are in relevant models (i.e. hSSD and Web-ICE) and of consequences for HC5 uncertainty.
	Validation of hSSD scenario-specific HC5s relative to the field and/or mesocosm studies.
	Critically review whether any of the growing amount of information types about chemicals and their impacts that is now available should be used to inform SSD development, application, and interpretation, including for example knowledge of omics, mechanisms, chemical properties, and exposure scenarios.
Toxicity data	Research is needed to determine how best to use available data (e.g. strict standardization criteria with resulting loss of species diversity or use weighting based on data quality). The focus of SSD development has been on acute toxicity data, and chronic toxicity estimation approaches will need the same level of evaluation (e.g. minimum data sets, acute to chronic ratio estimation, lowest toxicity value approaches). Develop better application of toxicological data in SSDs, e.g. using more chronic data, mechanistic understanding. Develop methods to expand on data availability by adding less strictly selected input data and putting less weight on their inclusion, based on reliability of data.
	Develop methodology to improve the use of predictive modelling to overcome limited data sets. The applicability of toxicity extrapolation method should be further validated for acute effects, and should also be evaluated for chronic effects. Develop and extend software tools to add the capacity to predict chronic toxicity and approaches applicable to other environmental compartments (such as sediment, soil and air) both remain significant research needs.
	Investigate the value of including microorganisms in SSDs to protect ecosystem functions e.g. when assessing the ecological risk of fungicides, investigate the effects of including various fungal species in the test battery and incorporating their data into the SSD; Microorganisms should be considered in the HCx derivation but development is currently hindered by the lack of available approved testing procedures for different groups of microorganisms.
Mechanistic understanding	Investigate whether critical body residue (CBR)-based SSDs could be developed.
	MOA (mode of action) is an important determinant of species sensitivity. Research is needed to determine linkages between MOA and SSD composition requirements. Investigate whether it is possible to treat MoA in the statistical models in the same way taxonomic distance is being used? (In particular, is this feasible for Web-ICE and hSSD?)
Uncertainty	Develop an understanding of uncertainties within the assessment that are currently unquantifiable. Studies should be conducted to identify the magnitude of the uncertainty of various components of the SSD methodology. Uncertainty may be related to lack of data, (non)representativity of data, mode of action considerations, and many other aspects of real exposure situations. An understanding of the mathematical magnitude of uncertainty alone may not be enough as it is possible that large sources of error may have little ecological importance, and vice-versa. Research should then be focused on reducing the uncertainty of the most important sources uncertainty in the SSD methodology. The group felt that uncertainty-driven research would be an important means to improve SSDs and maximize their usefulness in a cost-efficient manner. An uncertainty driven research agenda is also likely to increase uptake of the other methods that can be used in combination with SSDs e.g. QSARs, Web-ICE.
	A simple example of uncertainty-driven research would be the selection of chemicals (or species) to be used in

	ecotoxicity tests. If the toxicity of a chemical to a large number of species belonging to different taxonomic groups has been determined then the need for further research for that chemical may be low compared to a chemical that has been the subject of no or minimal toxicity testing. Another example is that very few SSDs have been conducted for non-chemical stressors (e.g. temperature, salinity) or the combined action of chemical and non-chemical stressors. Conducting such research could dramatically reduce uncertainty in the ecological relevance of single chemical SSDs, and place the risks posed by chemicals into a more meaningful context that addresses all possible pressures.