1

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§§

Running Head: Advancing Species Sensitivity Distributions in ecotoxicology

Key Words: Species Sensitivity Distribution, threshold, community, impact

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

18   † Procter & Gamble, Mason Business Center, P.O. Box 1093, Mason, OH 45040

19   ‡ U.S. Environmental Protection Agency, Gulf Ecology Division, 1 Sabine Island Drive, Gulf

20   Breeze, FL USA

21   § Durham University Department of Mathematical Sciences, South Road, Durham DH1 3LE,

22   UK

23 || ECETOC, Avenue E. Van Nieuwenhuyse, 2 B-1160, Brussels, Belgium

24   # Syngenta, Jealott's Hill Research Station, Bracknell RG42 6EY, UK

25   †† Safety & Environmental Assurance Centre, Unilever Colworth, Sharnbrook, Bedford, MK44

26   1LQ, UK

27   ‡‡ National Institute for Public Health and the Environment, Centre for Sustainability,

28   Environment and Health, P.O.Box 1, 3720 BA Bilthoven, the Netherlands

29   §§ Environment Agency, Red Kite House, Howbery Park, Wallingford, Oxon OX10 8BD, UK

ABSRACT

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 in decision support 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.

EDITORS NOTE:

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: (1) to

53   study and where possible improve the ecological relevance of SSDs, (2) to collate, compare and

54   where possible improve statistical approaches for SSD modeling, and (3) to describe and

55   evaluate regulatory applications of SSDs.

INTRODUCTION

Chemicals are an integral element of human society and their production, use, and potentially 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 relying on application factors (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;

79 Management of chemicals in the environment usually includes comparison of expected

80 exposures to a critical effect limit such as a Predicted No Effect Concentration for ecosystems

81 (PNEC) (ECHA 2008). Concentrations below the PNEC are considered to have a negligible

82 potential effect on the structure or function of an exposed ecosystem. When sufficient data are

83 available a PNEC may be estimated as a low percentile of an SSD (Van Straalen and Denneman

84 1989). PNECs are most commonly deterministic and estimated from applying an Application

85 Factor to the data derived from the most sensitive species tested (the actual AF being a function

86 of the type of data, acute or chronic, and the number of species tested). When PNECs are

87 estimated using SSDs, the extrapolation of laboratory test results to protect field populations and

88 communities results usually employs lower AFs (generally 1 to 5), while being somewhat

89 flexible to account for the biological diversity present in and the statistical qualities of the SSD

90 being considered (ECHA 2008).

91   Species Sensitivity Distributions have an established role in the assessment and

92 management of risks posed by chemicals, and major developments around the world have

93 provided relevant novel insights into their development and application. The formal adoption of

94 SSDs for the derivation of environmental thresholds dates back to scientific- and policy

95 milestones of 1985 in the United States and 1989 in Europe (Stephan et al. 1985; Van Straalen

96 and Denneman 1989). In 2001, SSDs were evaluated intensively for the derivation of European

97 environmental quality standards (EC 2001) and in 2002 the last comprehensive overview of the

98 principles and practices of SSD use on an international basis was made (Posthuma et al. 2002).

99   Here we summarize the major findings of a workshop sponsored by the European Centre

100 for Ecotoxicology and Toxicology of Chemicals (ECETOC) and the UK Environment Agency

101 held in February 2014 in Amsterdam, The Netherlands (ECETOC 2014). Forty experts from

102   academia, business, and government reviewed the state of the science for estimating toxicity

103   thresholds for aquatic ecological communities using SSD modeling, and considered advances in

104   statistical, ecotoxicological, and ecological science applicable to SSDs that have occurred since a

105   similar workshop was held in London in 2001 (EC 2001). New approaches or refinements to

106   current applications of SSD modeling were evaluated against current methods in which SSDs are

107   used in the context of environmental protection and management. The aim of this paper is to

108   provide an overview of the findings, conclusions and recommendations of the workshop.

109   **DERIVATION OF SPECIES SENSITIVITY DISTRIBUTIONS**

110   Predictive risk and retrospective impact assessment of chemicals requires estimation of

111   the toxicity thresholds of chemicals for aquatic communities as an integral aspect of defining

112   environmental hazard. Of the available tools used in hazard and risk assessment, SSDs provide a

113   particularly informative approach because they explicitly relate the intensity of chemical

114   pressure (e.g., the concentration) to ecological impacts (the proportion of species at risk).

115   Currently, hazard is most frequently predicted using concentration–effect data from single

116   species laboratory toxicity tests that measure effects on individuals and populations. Typically,

117   responses of individuals include survival (applicable to acute and chronic testing), growth and

118   reproduction endpoints for invertebrates, fish, amphibians and macrophytes.   Population

119   responses such as growth rate for microinvertebrates, bacteria, algal and cyanobacteria tests are

120   used in acute and chronic exposures. However, protection goals are generally broader than those

121   covered by endpoints derived from laboratory toxicity testing and focus on populations,

122   communities, and ecosystems. There is growing interest in moving from hazard levels derived

123   from individual toxicity tests to the use of SSDs, which can better be used to estimate potential

124   hazards to communities. Note that while SSDs include multiple species, they are the compilation

125   of individual species responses and typically do not include inter-specific interactions (predation,

126   competition) or ecosystems processes (nutrient cycling, energy flow).

127   The statistical methods and underlying scientific foundation supporting the use of SSD

128   models and the versatile use of these in environmental protection, assessment and management

129   were reviewed by Posthuma as discussed at the Workshop and reported earlier (ECETOC 2014).

130   Briefly, the SSD method assembles single species toxicity data to predict a hazardous

131   concentration (HCp) affecting a certain percentage (p) of all the species in a distribution, or to

132   estimate the toxic pressure, expressed as the potentially affected fraction (PAF) of species,

133   exerted on an assemblage from an observed or expected exposure concentration. SSDs can be

134   constructed using either acute or chronic test data, depending on data availability and they can be

135   related to the protection goal. In comparisons amongst chemicals, SSDs derived from ecotoxicity

136   data can have different positions (intercept) and shapes (slope) used to derive the HCp. The

137   higher the HCp of a chemical, the lower is its ecotoxic potential to induce impacts. Greater toxic

138   pressure is indicated by a larger PAF for a contaminated sample. The potential for expected

139   impacts for tested species and impacts on aquatic communities is therefore assumed to be

140   greater.

141   SSDs are constructed with the aim of predicting acute or chronic toxicity, although these

142   are usually dealt with separately. Single species data for acute toxicity (expressed as median

143   lethal or effective concentrations [LC50, EC50]), or estimates for chronic effects (expressed as,

144   no-observed-effect concentrations [NOECs], Chronic Values [defined as the geometric mean of

145   the NOEC and LOEC, Lowest Observed Effect Concentration], and EC10s) for several species

146   are fitted to one or more cumulative distribution functions followed by evaluation and choice of

147   the best model The cumulative distribution function is often assumed to be lognormal or log-

148 logistic (Awkerman et al. 2013; Posthuma et al. 2002). Other distributions have been used and

149 can also have utility (ANZECC and ARMCANZ 2000; Warne et al. 2015). A typical approach

150 uses the 5th percentile of the distribution of acute or chronic effects to derive toxicity thresholds

151 or environmental quality criteria that should ensure that the specified level of protection is

152 achieved. The estimation of the toxic pressure (PAF of species) given an exposure allows the use

153 of any endpoint (e.g., NOEC, EC10, EC50, LC50), depending on the expected level and duration

154 of exposure. Similarly, the estimated toxic pressure can yield assessment outcomes such as a

155 PAF_{NOEC}, or a PAF_{EC50} that specify the fraction of species exposed above their NOEC or EC50,

156 respectively. For example, an ambient exposure might predict that 50% of the species are

157 exposed above their NOEC whilst at the same time 20% of species are exposed above their

158 EC50.

159 One of the principle advantages of probabilistic SSDs over deterministic application

160 factors is the opportunity to express uncertainty in the point estimate (HCp) as additional

161 information for the risk assessor to judge the utility of the estimated threshold. Typically, the

162 HC5 will be accompanied by a confidence limit that conveys knowledge of the shape of the

163 statistical distribution of toxicity values and their variance. By addressing critical data that

164 appear to strongly influence the shape of the distribution (often at the tails of tolerance and

165 sensitivity) the risk assessor can understand the impact of particular data on the HCp and the

166 confidence interval around it.

167 **ECOLOGICAL, STATISTICAL, AND REGULATORY CONSIDERATIONS**

168 Since the sensitivity of all the species that might be exposed to a chemical cannot be

169 known, extrapolation needs to be done from the data available. ECETOC (2014) discussed that

170 scientifically sound extrapolation approaches based on SSDs to derive toxicity threshold

171 concentrations should provide a more useful and transparent assessment of risks than a

172 deterministic approach using generic factors applied to single species aquatic toxicity test data.

173 The SSD methodology is a valuable regulatory and management tool since it can provide greater

174 insight into the potential effects of a particular level of exposure compared to the deterministic

175 application factor method, enabling better problem definitions and decision support.

176   Regulatory tools such as SSD modelling are useful if they strike a balance between being

177 overcautious and under-protective. Being overly protective can lead to unnecessary mitigation

178 costs and stifle innovation whereas under protection may result in environmental degradation

179 (ECETOC 2014). A prospective risk assessment conducted in the context of environmental

180 protection needs to establish that there will be acceptable risk at the criterion concentration (e.g.,

181 Predicted No Effect Concentration for Ecosystems [PNEC], Environmental Quality Standard

182 [EQS], or Regulatory Acceptable Concentration [RAC]). In contrast, retrospective impact

183 assessment uses diagnostic tools to identify the cause of existing adverse effects, using SSDs to

184 quantify expected chemical impacts compared to other stressors (De Zwart et al. 2006). When

185 sufficiently large datasets are available, the risk of errors is reduced, while uncertainty on

186 expected protection or impact prediction declines. In such cases, SSD modelling provides a

187 mechanism for quantifying the relationship between chemical pressure and impact that takes

188 account of uncertainty due to differences in sensitivity between species. When datasets are small,

189 uncertainty is greater and consequently the more cautious deterministic approach may be more

190 appropriate. That is, the criterion is derived from the available data combined with an application

191 factor. Under conditions of small data sets (e.g., few species tested) or lower data quality, a

192 higher application factor is implied and appropriate for the deterministic assessment. Similarly,

193   the size of an assessment factor applied to an SSD will vary (minimum of 1) according to the

194   uncertainty in the hazard estimation.

195   Requirements for consideration of an SSD approach vary across regulatory jurisdictions

196   (e.g., by national regulatory authority), regulatory frameworks for specific compound classes

197   (e.g., pesticides covered under US FIFRA or EU PPP D [1107/2009]) or intended use in an

198   assessment framework (e.g., water quality standards or chemical-specific risk assessments).

199   Table 1 provides an overview of representative (not exhaustive) considerations in several

200   frameworks. It is interesting to note the variation in species coverage, treatment of multiple data

201   on the same species used as SSD input, and application of statistical principles that are applied.

202   The most recent guidances on SSD use for assessing hazards of chemicals (ECHA 2008) and

203   plant protection products (EFSA 2013) are not surprisingly the most complete across all the

204   facets to be considered. These guidances are consistent with discussions in Europe in the

205   previous decade (EC 2001; ECETOC 2008) and form the basis of subsequent national and

206   international guidance used in setting water quality criteria as well (e.g., CCME 2007; EC 2011).

207   ECETOC (2014) cautioned that continued validation of predictions made using SSDs

208   against a reference tier, such as field and mesocosm data, is required to ensure that a threshold

209   derived from an HCp (sometimes coupled with an application or safety factor) or a PNEC

210   (Predicted No Effect Concentration) has ecological relevance (see also Versteeg et al. 1999;

211   Posthuma et al. 2002). A new development is the advent of the SSD approach applied to field

212   data rather than field data being regarded as a separate line of evidence (Kwok et al. 2008). The

213   results of any extrapolation process (including SSDs) should always be critically assessed based

214   on all available knowledge on the substance and related substances, such as their mode of action

215   and other lines of evidence including field and mesocosm data. Use of the SSD methodology

216 should yield more generally conservative estimations of hazard (i.e., lower predicted effect

217 concentrations) and thus more readily acceptable results in most regulatory contexts than those

218 obtained from mesocosm-based methods (Versteeg et al. 1999). Differences remain across

219 regulatory jurisdictions on this aspect (for example, Canadian and Australian regulatory

220 decisions would place increased emphasis on mesocosm results if conducted following sound

221 statistical, biological and ecological principles; ANZECC 2000; CCME 2007). Mesocosms and

222 field studies will remain valuable tools for evaluating the accuracy of SSD predictions because of

223 the inherent interactions among populations and communities that are not inherent in single

224 species tests. Further, as acknowledged in many other venues, mesocosms often have the

225 additional advantage of utilizing more realistic field exposures (Giddings et al. 2002).

226   A new development in the use of SSDs is an emerging interest in using field data based

227 on population abundance and biomass as alternatives to toxicity estimates in the laboratory

228 (Leung et al. 2005). Field-based SSDs may allow an expansion of taxonomic coverage and thus

229 provide insight into responses for taxa less easily tested in the laboratory but that exist

230 temporally in the same space. On the other hand, intra- and inter-specific interactions as well as

231 multiple-stress responses are certainly involved in field assessments. Therefore, the interpretation

232 or meaning of the SSD may change compared with assessments based solely on laboratory single

233 species toxicity tests.

234   Multiple statistical approaches are available for SSD modeling and high uncertainty can

235 arise in cases of limited taxa diversity (ECETOC 2014). To address data gaps in taxa diversity,

236 the hierarchical SSD (hSSD) was developed as a novel approach and discussed by Craig and

237 colleagues (Craig et al. 2012; Craig 2013; ECETOC 2014). This can be used to predict

238 thresholds for defined species assemblages using knowledge of the general trends in how species

239 sensitivity is related to their taxonomic distance. Other methods for addressing data gaps in taxa

240 diversity include the U.S. EPA Web-ICE tool (www.epa.gov/ceampubl/fchain/webice/) which

241 uses interspecies correlation estimation models to estimate toxicity for taxa with limited data

242 (Awkerman et al. 2013). The U.S. EPA Web-ICE tool also explored interspecies toxicity

243 estimation as a function of taxonomic distance and showed the phenomenon is generally

244 important. While the investigations do not aim to assess the influence of chemical class on the

245 relationship, the fact that many modes of action are present in the database suggest it is a

246 generalized phenomenon. Traditional statistical approaches, Web-ICE, and the hSSD prototype

247 were compared and contrasted in ECETOC (2014) using case studies involving the surfactant

248 linear alkylbenzene sulfonate and the insecticide chlorpyrifos. Three distinct regulatory

249 applications associated with the use of SSDs are evident:

250 1. The derivation of generic protective threshold concentrations applied to many different

251   locations, perhaps over very large geographical regions. These are assumed to offer

252   sufficient protection everywhere, even in the most sensitive systems.

253 2. The derivation of scenario-specific protective thresholds that more closely reflect local

254   conditions (e.g., constrained to resident species or for a certain water quality condition),

255   but which may not be transferable from one place to another.

256 3. Identifying the causes of biological impact ('diagnosis') or expected impact magnitudes

257   of existing or expected (mixture) contamination, in order to inform the need and focus for

258   any remedial or management action.

259 The first 2 applications are protective and thus will tend to include a certain amount of

260 precaution, while in contrast the third needs to be predictive.

RESEARCH NEEDS

261

262   The overview of SSD practices as discussed during the workshop has shown that SSDs

263   currently have a significant influence on national and international decision making regarding

264   assessments of chemical exposure to ecosystems. It is evident from review of current

265   applications of SSDs in regulatory decision-making that better understanding of the state of the

266   science and answers to frequently asked questions would encourage best practices in the use of

267   SSDs by regulators, risk assessors, and risk managers. Although expert judgement has a role in

268   the interpretation of SSD models, a compilation of current best practices would provide a

269   valuable compendium of regulatory experiences beneficial to countries seeking to derive their

270   own environmental quality standards or to scientists seeking to understand the significance of

271   emerging chemicals or new applications of existing chemicals on ecosystems. An array of

272   modelling tools has extended the statistical evaluation of SSD "quality" that builds upon

273   progressively better and more available input data as a result of global chemical management

274   programs (e.g., OECD HPV [High Production Volume] Challenge program, European REACH,

275   Canadian Categorization of the Domestic Substances List and others). According to ECETOC

276   (2014) the use of species sensitivity distributions in ecological diagnostics links policy targets on

277   ecological integrity, monitoring data, SSD modeling and landscape-level mixture impact

278   diagnosis. Therefore, research that builds a stronger scientific foundation is preferable to work

279   focused narrowly on a single species or taxa.

280   Specific research needs were identified in the workshop that would augment the application of

281   SSDs in most circumstances: The research needs were divided into the following themes: use of

282   the SSD, ecological considerations, guideline considerations, model development and validation,

283 toxicity data, mechanistic approaches, and uncertainty (Table 2). The most important of these

284 are highlighted here.

285   1. Tools for regulatory decision making should be given high priority with particular focus

286   on i) SSDs for chronic toxicity, ii) validating HC5s with mesocosms and real ecosystems,

287   and iii) maximising the use of available data, e.g. by applying weighting criteria.

288   Rationale: the most potentially influential use of SSDs is establishing safe concentrations

289   for ecosystems associated with long term, low level exposure to chemicals, therefore

290   assessments based on chronic exposures are essential. However, the use of SSDs in

291   general should be somewhat more conservative (i.e., predict lower hazardous

292   concentrations) for routine use than higher tier studies (e.g., mesocosms). Higher tier

293   studies should still behave consistently with predictions provided by SSDs (Versteeg et

294   al. 1999). Acute SSDs also have a role and may be critical in some situations such as

295   short term pesticide exposures.

296   2. Mechanisms to maximize the use of available data should be further developed, e.g. by

297   applying weighting criteria to broaden taxonomic coverage and use of non-GLP (Good

298   Laboratory Practice) studies.

299   Rationale: The majority of standardized toxicity tests focus on relatively few species.

300   Taxonomic coverage is a key facet of developing SSDs and non-standard tests are

301   increasingly used as input. These are also most often not performed under a GLP

302   framework. Weighting or valuing different types of studies should be explored to

303   maximize the use of all high quality data that are available.

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

305   Rationale: Aquatic and sediment environmental exposures are rarely to single chemical

306   or stressor insults and are more commonly to mixtures. Methods to perform aggregate

307   and cumulative assessments are needed for the future as mixture assessments are

308   increasingly demanded by the stakeholders. Effluent toxicity assessments address this to

309   a degree but SSD-based mixture assessments are possible if mode of action and theories

310   of concentration addition and independent action can be accounted for (Kapo et al. 2014).

311   4. Trait-based SSDs appear to offer advantages over conventional taxonomic based

312   approaches, but there is currently no practical application.

313   Rationale: This continues to be a developing science in ecotoxicology. It is likely that

314   responses to chemicals are in part based on ecological traits (much like their

315   classifications in feeding or trophic ecology) with some trait types more sensitive to

316   certain types of exposures than others (Pilière et al. 2014).

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

318   Rationale: It is well established that photosynthetic micro-algae are frequently more

319   sensitive than fish or invertebrates (Jeram et al. 2005) but are sometimes not considered

320   in SSD formulation. Photosynthetic and non-photosynthetic microbes, aquatic

321   macrophytes and plants play crucial roles in ecosystem structure and function, therefore,

322   including these species in SSDs more frequently may improve robustness of predictions.

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

324   Rationale: Future environmental toxicology approaches should be able to take advantage

325   of the large efforts on-going in efforts such as the US NRC "Toxicity Testing in the 21st

326   Century" (NRC 2007).   Cheminformatics is the strategic use of computer and

327   informational techniques applied to a range of problems in the field of chemistry

328   including those of drug discovery, development of in silico models, and relating key

329   chemical attributes to the potential for hazard. Environmental scientists generally have a

330   strong appreciation for physical-chemical attributes in testing and assessment that will

331   bridge well to cheminformatics. How SSDs approaches can take advantage of this will

332   be explored.

333   7. Focus on sensitive groups.

334   Rationale: A better understanding of the frequency of bi-modality in SSDs is needed (i.e.,

335   when one taxonomic group is more sensitive compared to others) and how to further

336   incorporate this into assessment methodologies is needed. Certain groups of chemicals

337   may even benefit from a greater focus on sensitive subgroups, for example micro-algae to

338   anti-microbials, as a stronger basis for extrapolation for environmental protection.

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

340   Rationale: Approaches of the h-SSD form provide some unique advantages to probe

341   relationships between available studies used as SSD inputs and actual distributions of

342   species based on taxonomy observed in the field (Craig et al. 2012; Craig 2013).

343   9. Internal dose (CBB or critical body burden)-based approaches have potential to

344   incorporate mechanistic toxicokinetic/toxicodynamic modelling approaches that could

345   help explain sensitivity differences between taxa/traits.

346   Rationale: Critical body burden concepts allow a technically defensible determination of

347   exposure to chemicals at the target organ of interest resulting in acutely or chronically

348   toxic effects (McCarty et al. 1992; McElroy et al. 2011). CBB approaches have generally

349   been investigated for organic compounds and are not only more mechanistically-based, a

350   laudable goal in any toxicological investigation, but also have the attractive feature of

351   providing insight into mixture assessments. Greater emphasis on developing CBB for

352   algae and invertebrates would need to be undertaken as fish have been the primary group

353   of interest until now. This also highlights the potential for various modes of action being

354   appropriate for a single chemical, e.g., in different species possessing different

355   physiologies, traits and responses.

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

357   Rationale: It is acknowledged that this will be a challenge for any regulatory framework;

358   however, it is consistent with the goals of risk assessment which is fundamentally

359   probabilistic in nature. Research is needed to ascertain the relationship of statistical

360   uncertainty with deterministic application factors typically applied to small data sets.

361   Improvements to the role of application factors, even as they are applied to SSD results,

362   due to variation in SSD quality, are also warranted.

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

364   Rationale: Through the development of more unified global best practices, the means to

365   value the varying levels of quality resulting from SSD methods may become clear.

366   Treatment of data (multiple studies on the same species, different endpoints utilized even

367   for the same species), taxonomic coverage (breadth of species, species choices),

368   statistical models used, and how these affect HC5 predictions and their uncertainties is

369   essential for long term support of the tool.

376   **REFERENCES**

377   ANZECC (Australian and New Zealand Environment and Conservation Council) and

378   ARMCANZ (Agriculture and Resource Management Council of Australia and New

379   Zealand). 2000. Australian and New Zealand Guidelines for Fresh and Marine Water

380   Quality Volume 2 Aquatic Ecosystems — Rationale and Background Information (Chapter

381   8). 678 pp.

382   Awkerman JS, Raimondo S, Jackson CR, Barron MG. 2013. Augmenting aquatic species

383   sensitivity distributions with interspecies toxicity estimation models. *Environ Toxicol*

384   *Chem* 33:688-695.

385   CCME (Canadian Council of Environmental Ministers). 2007. A Protocol for the Derivation of

386   Water Quality Guidelines for the Protection of Aquatic Life (Draft March 14, 2007). Part

387   1. Canadian Water Quality Guideline Framework. Ottawa, Canada 36pp.

388   Craig PS, Hickey GL, Luttik R, Hart A. 2012. Species non-exchangeability in probabilistic

389   ecotoxicological risk assessment. *J Royal Stat Soc. Series A: Statistics in Society* 175:243-

390   262.

391   Craig PS. 2013. Exploring Novel Ways of Using Species Sensitivity Distributions to Establish

392   PNECs for Industrial Chemicals: Final Report to Project Steering Group. Downloaded

393   from: http://dro.dur.ac.uk/13383 (last accessed 10/02/2016)

394   De Zwart D, Dyer SD, Posthuma L, Hawkins CP. 2006. Predictive models attribute effects on

395   fish assemblages to toxicity and habitat alteration. *Ecol App* 16:1295-1310.

396   EC (European Commission). 2001. EC Expert Consultation Workshop on Statistical

397   Extrapolation Techniques for Environmental Effects Assessments. Report of the London

398   Workshop, 17-18th January, 2001; European Commission, Joint Research Center, Institute

399   for Health and Consumer Protection, European Chemicals Bureau, Ispra (VA), Italy. 9 p.

400   EC (European Commission). 2011. Technical Guidance for Deriving Environmental Quality

401   Standards. CIS Guidance Document No. 27, Technical Report – 2011 – 055.

402   ECETOC (European Center for Toxicology and Ecotoxicology of Chemicals). 2008. Workshop

403   on the probabilistic approaches for marine hazard assessment, 18-19 June 2008, Oslo,

404   Norway. Brussels, Belgium. Workshop Report No. 15, 83 pp.

405   ECETOC (European Center for Toxicology and Ecotoxicology of Chemicals). 2014. Estimating

406   toxicity thresholds for aquatic ecological communities from sensitivity distributions. 11-13

407   February 2014, Amsterdam. Workshop Report No. 28. European Centre for Ecotoxicology

408   and Toxicology of Chemicals. Brussels, Belgium.

409   ECHA (European Chemicals Agency). 2008. Guidance on information requirements and

410   chemical safety assessment Chapter R.10: Characterization of dose [exposure]-response for

411   environment. May, 2008. 65 p. Downloaded from: http://echa.europa.eu/reach_en.asp.

412   EFSA (European Food Safety Authority). 2013. Scientific Opinion, Guidance on tiered risk

413   assessment for plant protection products for aquatic organisms in edge-of-field surface

414   waters, EFSA Panel on Plant Protection Products and their Residues (PPR). EFSA Journal

415   2013 11(7):3290, 268p.

416   Giddings JM, Brock TCM, Heger W, Heimbach F, Maund SJ, Norman SM, Ratte HTA,

417   Schaefers C, Streloke M. 2002. Community-Level Aquatic System Studies –

418   Interpretation Criteria. Society of Environmental Toxicology and Chemistry, SETAC

419   Press. 60p.

420   Jeram S, Sintes JM, Halder M, Fentanes JB, Sokull-Klüttgen B, Hutchinson TH. 2005. A

421   strategy to reduce the use of fish in acute ecotoxicity testing of new chemical substances

422   notified in the European Union. *Regul Toxicol Pharmacol* 42: 218-24.

423   Kapo KE, Holmes CM, Dyer SD, De Zwart D, Posthuma L. 2014. Developing a foundation for

424   eco-epidemiological assessment of aquatic ecological status over large geographic regions

425   utilizing existing data resources and models. *Environ Toxicol Chem* 33:1665-1677.

426   Kwok, KWH, Bjorgesæter A, Leung KMY, Lui GCS, Gray JS, Shin PKS, Lam PKS. 2008.

427   Deriving site-specific sediment quality guidelines for Hong Kong marine environments

428   using field-based species sensitivity distributions. *Environ Toxicol Chem* 27:226-234.

429 Leung KMY, Bjorgesater A, Gray JS, Li WK, Lui GCS, Wang Y, Lam PKS. 2005. Deriving

430   sediment quality guidelines from field-based species sensitivity distributions. *Environ Sci*

431   *Technol* 39:5148-5156.

432 McCarty, LS, Mackay D, Smith AD, Ozburn GW, Dixon DG. 1992. Residue-based

433   interpretation of toxicity and bioconcentration QSARs from aquatic bioassays: Neutral

434   narcotic organics. *Environ Toxicol Chem* 11:917-930.

435 McElroy AE, Barron MG, Beckvar N, Kane SB, Driscoll T, Meador JP, Parkerton TF, Preuss

436   TG, Stevens JA. 2011. A review of the tissue residue approach for organic and

437   organometallic compounds in aquatic organisms. *Integ Environ Assess Manag* 7:50-74.

438 NRC (National Research Council). 2007. Toxicity testing in the 21st century : a vision and a

439   strategy. National Academies Press, Washington, DC.

440 OECD (Organization for Economic Cooperation and Development). 1992. Report of the OECD

441   Workshop on Extrapolation of Laboratory Aquatic Toxicity Data to the Real Environment.

442   OECD Monograph No. 59. Paris, France. 45p.

443 Pilière A, Schipper AM, Breure, AM, Posthuma L, De Zwart D, Dyer SD, Huijbregts MAJ.

444   2014. Comparing responses of freshwater fish and invertebrate community integrity along

445   multiple environmental gradients. *Ecolog Indicators* 43:215-226.

446 Posthuma L, Suter GW, Traas TP. 2002. Species sensitivity distributions in ecotoxicology.

447   Lewis Publishers, CRC Press. Boca Raton, Florida.

448 Stephan CE, Mount DI, Hansen DJ, Gentile JH, Chapman GA, Brungs WA. 1985. Guidelines

449   for Deriving Numerical National Water Quality Criteria for the Protection of Aquatic

450   Organisms and Their Uses. National Technical Information Service, Washington D.C.

451   USA. 104 p.

452   UNEP (United Nations Environmental Programme). 2013. UNEP *Global chemicals outlook.*

453   *Towards sound management of chemicals*; United Nations Environment Programme:

454   Geneva, Switzerland. p. 248

455   Van Straalen NM, Denneman CAJ. 1989. Ecotoxicological evaluation of soil quality criteria.

456   *Ecotoxicol Environ Saf* 18:241-51.

457   Versteeg DJ, Belanger SE, Carr GJ. 1999. Understanding single-species and model ecosystem

458   sensitivity: data based comparison. *Environ Toxicol Chem* 18:1329-1346.

459   Warne MStJ, Batley GE, van Dam RA, Chapman JC, Fox DR, Hickey CW and Stauber JL.

460   2015. Revised Method for Deriving Australian and New Zealand Water Quality Guideline

461   Values for Toxicants. Department of Science, Information Technology, Innovation and the

462   Arts, Brisbane, Queensland. 50 pp.

463   Wheeler JR, Grist EPM, Leung KMY, Morritt D, Crane M. 2002. Species sensitivity

464   distributions: data and model choice. *Mar Poll Bull* 45:192-202.

1　Table 1. Examples of the use of Species Sensitivity Distributions in several regulatory frameworks for the purposes of chemical risk assessment and
2　formulation of water quality criteria or standards. The table is non-exhaustive and other frameworks are used by hazard assessors globally.
3

Facet of SSD Development	Factors to be considered	Standard for Acceptance in Regulatory Framework			
		ECHA 2008	USEPA Ambient Water Quality Criteria, Stephan et al. 1985	CCME 2013	ANZECC (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 relevant	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 estimates calculated at the	Median value for comparable records for the same endpoint when more than two data points are available; if only two the geometric mean is	Data for each species and endpoint expressed as a geometric mean and the lowest value per

			genus level as geometric means	used to represent the average species effects endpoint	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 lognormal and log logistic identified as most common	Log triangular required applied to the four most sensitive genera	Confirm model choice	Burr Type III distribution recommended (note this includes log-normal, log-logistic 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

4

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

2 practitioners.

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.