Paparella et al.:

Limitations and Uncertainties of Acute Fish Toxicity Assessments Can Be Reduced Using Alternative Methods

Supplementary Data

Tab. S1: Characterization of the acute fish toxicity test (OECD TG 203)
According to the OECD template for individual information sources to be used within IATAs (OECD, 2016; gray shaded and non-shaded lines), supplemented with information on uncertainties related to each characterization element (red shaded lines) and how alternative methods or approaches may reduce these uncertainties (green shaded lines). Cross references are indicated in blue.

	SCIENTIFIC BASIS/RELEVANCE, INCLUDING MODE OF ACTION COVERAGE	
1	Lethality, measured as an indicator of potential population decrease.	
1.1	Underlined: Uncertainties: Lethality under laboratory conditions is a crude indicator for potential population decrease. The moribundity criterion may be environmentally more relevant and more protective. This hypothesis appears mechanistically plausible, since a weakened fish is likely to be a dead fish in the real-world environment due to predators, competitors and/or other environmental stressors (Rufli, 2012)[1]. However, using moribundity as the endpoint rather than lethality was not agreed by the OECD member countries and was not included in the 2019 update of TG 203 (see 2.3 and 4.1).	
1.2	Alternative methods may facilitate the assessment of sub-lethal endpoints, which could represent, mechanistic indicators of an increased probability of environmental population level lethality. Endpoints in the Fish Embryo Acute Toxicity Test (FET) (coagulation of fertilized eggs, lack of somite formation, lack of detachment of the tail-bud from the yolk sac and lack of heartbeat) are mechanistic in that they provide more information than simply whether the fish is dead/alive. If needed, these endpoints could also be expanded to include others such as neurotoxicity (Stengel et al., 2018; Zindler et al., 2019; Kluver et al., 2015). Based on the hypothesis that acute fish toxicity is often due to nonspecific modes of action, endpoints such as metabolic activity and cell- and lysosomal membrane integrity measured in the RTgill W1 cell-line test (Fischer et al., 2019) could be considered as a mechanistic key event, even without a fully characterized Adverse Outcome Pathway (Volz et al., 2011)[2]. Such effects on rainbow trout gill cells – similar to sublethal endpoints in the FET – are likely to increase the likelihood for environmental population level lethality. This hypothesis appears mechanistically plausible, because compromised or weakened fish are likely to die in the real-world environment due to predators, competitors and/or other environmental stressors[1]. If needed, also other specific endpoints could be tested *in vitro*, such as acetylcholinesterase-inhibition (Arini et al., 2017)[3]. Computational approaches, e.g., freeware such as OECD QSAR Toolbox[4], US EPA TEST[5], VEGA (Benfenati et al., 2013)[6] or commercial software, such as CATALOGIC[7] and iSafeRat[8] (Thomas et al., 2019) may inform on aquatic toxicity and structural alerts for non-baseline compounds and provide mechanistic information (Bauer et al., 2018a,b; Thomas et al., 2019). Such methods may be combined within an Integrated Approach to Testing and Assessment (IATA), which is work underway at OECD level (OECD WNT project 2.54). Computational approaches for data integration, e.g., Bayesian Networks (Lillicrap et al., 2020; Moe et al., 2020) could complement the IATA.	

[1] Appropriate data to support this hypothesis are not available for fish, but for algae and invertebrates (Zhao et al., 2020; Knillmann, et al., 2012).

[2] The key relevance of cytotoxicity as a mechanism for acute toxicity was also recognized for acute mammalian toxicity (Prieto et al., 2019; Vinken and Blaauboer, 2017).

[3] an AOP "Acetylcholinesterase inhibition leading to acute mortality" is under development see https://aopwiki.org

[4] The OECD QSAR Toolbox. Organisation for Economic Co-operation and Development. http://www.oecd.org/chemicalsafety/risk-assessment/oecd-qsar-toolbox.htm (updated 15.04.2020; accessed 15.05.2020).

[5] Toxicity Estimation Software Tool (TEST). US Environmental Protection Agency. https://www.epa.gov/chemical-research/toxicity-estimation-software-tool-test (updated 22.08.2016; accessed 15.05.2020).

[6] Virtual models for property evaluation of chemicals. Vegahub. https://www.vegahub.eu/ (accessed 15.05.2020).

[7] Software. Oasis Laboratory of Mathematical Chemistry. http://oasis-lmc.org/products/software/catalogic.aspx (accessed 15.05.2020).

[8] iSafeRat Online. KREATiS. https://isaferat.kreatis.eu/en/qsar.php (accessed 15.05.2020).

	It is noted that considering the environmental extrapolation uncertainty, the huge environmental variability and the uncertainty about this variability (see 2.1, 2.3, 4.3, 5.1, 6.1, 8.1, 9.1, 10), LC_{50} values from TG 203 should be considered as coarse indicators of potential population level lethality in real environments. Thus, it remains an open question, how precisely such coarse indicators should be predicted by other toxicity indicators from alternative methods.	
2	Fish are considered to be one of the highest trophic levels in ecotoxicology. Fish acute toxicity data are used together with acute invertebrate (daphnids) and algae data to inform on potential aquatic toxicity (in terms of these regulatory base-set requirements).	
2.1	Uncertainties: Information is available regarding the variability of LC_{50} values between the 6 fish species most frequently used in TG 203 tests (see 10), however, it remains uncertain how well this variability estimate represents the sensitivity distribution for all fish species and life stages in the environment. For 65 to 90% of chemicals, pesticides, and pharmaceuticals, daphnids and algae are more sensitive to toxicants than fish (Hutchinson et al., 2003; Jeram et al., 2005; Hoekzema et al., 2006; Rawlings et al., 2019). This range estimate is based on several analyses and thus indicates biological variability without a high uncertainty component. However, it indicates how often routine acute fish toxicity testing increases the protection level (compared to testing with daphnids and algae only). The range estimate also indicates that LC_{50} variability based on TG 203 studies, or estimated from alternative methods, may often not influence the PNEC derivation and GHS classification. Regarding the broader environmental protection goal, the use of TG 203 data causes numerous sources of extrapolation uncertainty relating to the complexity of natural environments. Water temperature, hardness, oxygen content, pH, sunlight and turbidity, water current, total organic carbon, oxidation, evaporation, presence of sediment, variable fish biotransformation and other aquatic degradation or transformation, and co-exposure with other stressors could impact on the toxicity of chemicals. Acute fish toxicity is estimated, maximizing the bioavailabilty of a single chemical for a single species, using standardized parameters not reflecting the potential natural environmental variability. Organism health, stress from variable food webs and other biotic factors may also influence the fish lethality in natural ecosystems. These enviornmental variables may increase or decrease the toxic potential of a compound relative to the experimental laboratory situation. Furthermore, the aquatic environment contains hundreds of thousands of species (Mora et al., 2011) and various life stages (embryonic, juvenile, adult, senescent). It is an extreme environmental simplification to cover this biodiversity by using short term tests for a specific life stage of three species, representing the three trophic levels. Data-based knowledge is available that single species tests are "*in a majority of cases, reliable qualitative (some level of response seen) predictors of aquatic ecosystem community effects*" (de-Vlaming and Norberg-King, 1999). This latter US-EPA review comprised 57 studies (74%) that support this conclusion and 16 studies (21%) where single species testing underestimated aquatic ecosystem effects, and 4 studies (5%) that were inconclusive. The review also explains that full quantitative validation of single species tests through field studies is not feasible and not meaningful given the substantial environmental variability. This also means that a single environmentally "true" value does not exist, regardless of the assessment method applied. (For quantitative extrapolation uncertainty, see 5 to 7.1).	
2.2	Informed species extrapolation is possible using computational methods, engaging for example, chemical structure, physicochemical properties, MoA classification, existing *in vivo* fish data, and/or other parameters (see 6.2). However, it is important to consider that models are only as good as the data used to build and train them. If poor quality, or highly variable AFT data is used in model construction, the models will suffer. Thus, models may not accurately predict experimental *in vivo* responses.	
2.3	Uncertainties: By default, TG 203 does not provide information on mechanism of toxicity. This limits *inter alia* extrapolation to other species and taxa (via Inter species Correlation Estimate (ICE) models and Ecological Threshold of Concern, EcoTTC; see 6.2). This extrapolation is more uncertain, without knowledge of the underlying mechanism and its relevance for the broader ecosystem. Introducing moribundity and eventually additional mechanistic endpoints by default into TG 203 would be theoretically feasible. As mentioned before, moribundity would also partly address the animal welfare concern. However, there is currently no agreement on the link between potential clinical signs and related mechanisms leading to mortality. Therefore, OECD VMG-Eco agreed to incite voluntary collection of clinical signs first and their relation to lethality at a later stage. The use of moribundity would not improve the standardization and the throughput for testing and assessment. Moreover, it would require the development of reference databases and validation (which would necessitate to continue animal testing). Considering the current validation stage and perspectives for alternative methods, it may be more efficient to invest into alternative methods based IATAs.	
2.4	Alternatives can provide some mechanistic information, see 1.2.	
DESCRIPTION		
3	At least 7 juvenile fish are exposed to the test chemical for 96 h under static, semi-static or flow-through conditions. Mortalities and visible abnormalities are recorded, and where possible the LC_{50} is determined. One of the following species are recommended: zebrafish, fathead minnow, common carp, Japanese medaka, guppy, bluegill, rainbow trout, three-spined stickleback, sheepshead minnow, European sea brass and sea bream. If other species are used, the rationale for the selection of the species must be reported together with any adaptation of the test guideline's recommendations.	

3.1	<u>Uncertainties for the selection of a specific test design (within the range of TG 203)</u> TG 203 stipulates that the selection of any fish species should depend on regulatory requirements and environmental exposure scenarios (cold, temperate or warm water species, freshwater or estuarine/marine fish), but does not provide further guidance for this species selection or discuss whether/when multiple species should be tested. TG 203 requires that juvenile fish are used. A range of acceptable fish-lengths for the different species is provided. However, because length is correlated as a cubic function to weight, small differences in length can still translate to larger differences in weight and developmental stage; see Table 6 in (Belanger et al., 2013). TG 203 does not specify the use of males or females. TG 203 does not precisely specify the number of fish to be used per concentration (see 5 and 5.1). This flexibility of the TG may introduce variability into the data, with disadvantages for global regulation (see 10, 10.1 and 16).
3.2	Alternative experimental methods are generally more standardized. They are not intended to provide an (uncertain) lethality estimate for a specific species and related environment. They are intended to provide a useful basis for estimating fish toxicity at least for the current standard species included in TG 203. For example, TG 236 (FET) limits variability of species to zebrafish, precisely defined developmental stages and related water parameters. The RTgill W1 fish cell line test (ISO, 2019) clearly defines the cell line, culture and exposure conditions. This may lead to significant advantages for regulatory toxicology, see 6.2, 10.2, 16.1.
RESPONSE MEASURED	
4	Lethality: Fish are considered dead, if there is no visible movement (e.g., gill movements) and if touching the caudal peduncle produces no reaction. Visible abnormalities are also observed and recorded, including loss of equilibrium, swimming behavior, respiratory function and pigmentation.
4.1	<u>Uncertainties:</u> In regulatory practice not lethality, but moribundity is legitimately[9] used as the endpoint in several countries, though this is not the case in all countries and TG 203 still prescribes lethality to be used. This introduces variability and uncertainty into the LC_{50} estimates. On the one hand, the use of moribundity may reduce the LC_{50} on average by a factor of 2, compared to lethality (Rufli, 2012). On the other hand, observations of moribundity are likely more subjective than observations of mortality and may introduce additional variability to the assay result. However, the variability and uncertainty from the use of moribundity or lethality can be estimated and reduced as soon as unambiguous criteria for moribundity have been agreed (Rufli, 2012). Thus, there is still uncertainty related to the use of these endpoints, but in principle the lethality to moribundity difference could be scientifically calibrated.
4.2	See 3.2.
4.3	<u>Uncertainties:</u> Since TG 203 is a vertebrate animal test, the basic test design has to be limited in terms of animal numbers. Therefore, there is no true biological replication in TG 203. All animals stem from one cohort and there are no tank replicates. Furthermore, TG 203 does not require positive controls, which would allow documenting the laboratories proficiency in acute fish toxicity testing, expectable intra- and inter-laboratory variability of LC_{50} values, within and between species and the validity of the individual study.
4.4	Alternative methods may include more replicates and include positive controls without significantly compromising 3R goals or increasing testing costs (especially if *in vitro* methods or computational methods are used, see 16.1).
PREDICTION OR EXTRAPOLATION MODEL	
5	The LC_{50} is calculated from the concentration-response relationship after 96 h. It is not mandatory to identify a concentration causing 0 or 100% mortality. The OECD TG 203 requires at least 7 fish to be used per concentration.
5.1	<u>Uncertainties:</u> The precision of experimental LC_{50} estimates is reduced when fewer than 20 fish are used per concentration, with flat dose-response relationships or in cases where the LC_{50} is off-center relative to the boundaries of the tested concentration range (Carr et al., 2018)[10]. A variety of models are used to estimate the LC_{50} which accommodate these difficult data structures and the LC_{50} estimate is sensitive to model selection in these cases. Confidence and prediction intervals are also subject to uncertainty when LC_{50} estimates from many tests, summarized by different methods, are being compared (Carr et al., 2018).
5.2	Alternative experimental methods may use more replicates and a larger number of test concentrations with no or less conflict with the 3Rs (e.g., fish cell lines (ISO, 2019) or TG 236, respectively).

---

[9] According to Directive 2010/63/EU, Article 13: "Death as the end-point of a procedure shall be avoided as far as possible and replaced by early and humane end-points".

[10] As extensively discussed in the related manuscript´s section 2, the authors do not advocate for an increase in the number animals used in TG 203 or for any other potential TG 203 improvements that would not advance the 3Rs, but to change towards the regulatory use of alternative methods.

	Computational models are built using a large amount of toxicological data for many chemical structures[11].
6	TG 203 LC_{50} may be used to derive an aquatic Predicted No-Effect Concentration (PNEC): current regulatory guidance suggests the use of default extrapolation factors (up to 1000) depending on the availability of just acute or also chronic toxicity data, from additional trophic levels (algae and/or daphnids) (ECHA, 2008).
6.1	Uncertainties: The variability of TG 203 derived LC_{50} values (see 10 to 10.1) may translate to variable PNEC extrapolations (in up to 35% of cases, see 2.1). Furthermore, in the absence of a standard validation of TG 203, there is uncertainty in this variability (see 9 and 10). Furthermore the assessment factors used according to current regulatory guidance are pragmatic, simplistic and result in PNECs, for which the environmental protection level (e.g., % species under risk) and related uncertainty is unknown (though there is some qualitative science basis for the regulatory use of AFT $LC_{50,}$ see 2.1 last paragraph). With regard to potential fish interspecies differences see 10. PNECs derived by expert risk assessors can vary by 3 orders of magnitude with the largest contributor being selection of the input data for PNEC derivation based on study quality (Hahn et al., 2014). When different expert groups may come to different, scientifically legitimate conclusions this represents ambiguity. It results from data complexity as one form of uncertainty. With increasing number of toxicological methods and approaches and the regulatory requirement for Weight of Evidence assessments, this type of uncertainty is likely to grow. Several tools including ICE, SSD, CTD and EcoTTCs models can be used to describe, model, and account for the variability across species. On this basis a single experimental fish LC50 (or prediction thereof) can be recognized as a very uncertain estimate for environmental aquatic toxicity. Confidence intervals for more comprehensively informed environmental toxicity estimates, such as the 5th percentile of SSDs may span 2 orders of magnitude (Awkerman et al., 2014; Bejarano et al., 2017). Moreover, the SSD models do not include the variability of fish toxicity due to the variability of environmental factors.
6.2	Variability for LC_{50} or EC_{50} estimates (and respective PNECs) from alternative methods may be lower (see 10.2) To counterbalance increase in complexity due to the regulatory requirement for Weight-of-Evidence assessments, alternative methods are usually intended to be used within standardized Integrated Approaches to Testing and Assessment (IATAs) and/or Integrated Testing Strategies (ITS) or Defined Approaches (DAs). The extrapolation-uncertainty from the use of pragmatic assessment factors would apply also to alternative experimental approaches. However, computational models may reduce the uncertainty of the current default assessment factor applications. Interspecies Correlation Estimations (ICEs; from species with test data to species without test data) may increase the number of chemicals with reliable Species-Sensitivity-Distributions (SSDs). The latter may inform on expected interspecies differences for chemical groups and inform on useful limit value derivations (Bejarano et al., 2017). Chemical Toxicity Distributions (CTDs) may inform on chemical-group specific expected lowest No Observed Adverse Effect Concentrations (NOAECs) and may build the basis for the development of an Ecological Threshold of Concern (EcoTTC) (Belanger et al., 2015). It is noted that the data used for model development stem from laboratory experiments and can never inform on the almost endless real environmental variability, but they allow best informed use of available knowledge.
7	The TG 203 $LC_{50,}$ in combination with algae and daphnids toxicity data and eventually information for biodegradability and bio-concentration, may be used to classify chemicals and mixtures 1) according to GHS[12] and 2) to define the T-criterion within the assessment of PBT (persistent, bio-accumulating, toxic) properties (LC_{50} < 0.01 mg/L; ECHA, 2017b), with potentially severe legal consequences, like the requirement to be banned and substituted on the market.
7.1	Uncertainty: There is variability in any type of experimental data. Therefore, especially for LC_{50} values close to classification cut off levels, the classification may change due to data variability, i.e., chance, rather than true differences in toxicity (Dimitrov et al., 2016). In the absence of agreed variability estimates for the acute fish, daphnia and algae data and the variability of the lowest LC_{50} or EC_{50} from such tests, there is uncertainty in this variability (see 9 and 10).
7.2	See 10.2.
METABOLIC COMPETENCE	

---

[11] QSAR predictivity may be improved by local validity assessment (Benfenati et al., 2013. VEGA-QSAR: AI inside a platform for predictive toxicology. CEUR Workshop Proceedings. http://ceur-ws.org/Vol-1107/) and/or developing QSARs specific for chemical groups sharing a molecular initiating event. Thus QSARs may also trigger investigation of experimental data, in case the latter deviate from high quality computational predictions (see, e.g., figure 1 in Thomas et al. (2019). How in silico and QSAR approaches can increase confidence in environmental hazard and risk assessment. *Integr Environ Assess Manag 15*, 40-50. doi:10.1002/ieam.4108).

[12] According to GHS and CLP regulation (EC No 1272/2008), LC_{50} values higher than 1 mg/L do not lead to acute aquatic toxicity classification whereas LC_{50} values ≤ 1 mg/L lead to classification in category 1. No other acute categories are defined, but further LC_{50} stratification into magnitudes (1-0.1 mg/L, 0.1-0.001 mg/L, etc.) allows the attribution of M-factors to the classified chemicals and this supports better mixture classification. Also classification for chronic toxicity categories is possible based on acute fish LC_{50} data, depending on information for ready biodegradability (ECHA, 2017c).

8	The metabolic biotransformation in juvenile fish can activate compounds resulting in enhanced toxicity when compared to the parent compounds. Alternatively, biotransformation could reduce the internal bioavailable concentration of the toxic compound and lead to a reduced toxicity.
8.1	Uncertainty: For compounds that require activation, differences in the LC_{50} of > 50-fold have been identified for different species (Scholz et al., 2016). Since TG 203 does not cover embryonic life stages, potential acute toxicity specific for this life stage might not be covered (e.g., life stage specific MoAs and/or absence/presence of specific biotransformation). There is a lack of knowledge about the biotransformation variability of biotransformation between the species recommended in TG 203 as well as a lack of knowledge about the variability of biotransformation between the different fish life stages. This has not been systematically assessed. Furthermore there is uncertainty with respect to modulations within real-world ecosystems (Schlenk et al., 2008). Therefore, a comparison between a juvenile fish and alternative test systems is difficult given the lack of information.
8.2	Biotransformation capacity can be well characterized for the validated alternative models, e.g., for the fish embryo acute toxicity test (Braunbeck et al., 2020; Tierbach et al., 2018) or for tests with fish-cell-lines (Stadnicka-Michalak et al., 2018). Thus the uncertainty about variability due to potentially variable biotransformation is reduced. This does not exclude the possibility that for some compounds a reduced toxicity may be observed due to a lack of bioactivation when using some alternative methods. So far, however, only one example (allyl alcohol) has been described as having a lower toxicity due to a lack of bioactivation when tested using TG 236 (Kluver et al., 2014; Knobel et al., 2012) or the RTgill-W1 fish cell line (Tanneberger et al., 2013) compared to TG 203. However, this may also apply when a certain single species is used for the AFT.
STATUS OF DEVELOPMENT, STANDARDIZATION, VALIDATION	
9	TG 203 was adopted by the OECD in 1981, and updated in 1983, 1992 and 2019. The most recent update is based on the review of the fish testing framework (OECD, 2012a) and includes: a) specifications, in terms of fixing the formerly flexible test duration and the maximum fish load for flow through conditions (g/L), guidance for the use of solvent and solvent-controls (OECD, 2019), guidance on analysis of the dilution water control and solvent control, guidance on the use the statistical methods, the fish length terminology, specifying suitable estuarine and marine species, specifying the use of juvenile fish (by species specific ranges for fish length), a 14 day interval between treatment against diseases and test initiation and indicating potential range finders. b) adaptations, in terms of deleting the need for a concentration range leading to a 0 to 100% lethality range and enhancing the recordings for visible abnormalities (for future use of more humane endpoints in line with (OECD, 2002). Some other elements recommended for review in (OECD, 2012a) could not be improved so far, such as the use of moribundity as a definitive indicator of lethality and as a trigger to terminate the experiment (see 1, 1.1 and 4.1), the lack of a positive control, lack of test-tank replicates (see 4.3), lack of guidance on fish-species selection among the 11 species recommended (see 3.1).
9.1	Uncertainties: There is some qualitative scientific basis for the regulatory use of AFT LC_{50} (see 2.1, last paragraph). However, no international official validation is available according to the present requirements for alternative methods, in terms of reliability and relevance for real aquatic environments. No validation is available with respect to the effect of environmental modifiers. With a view to analyzing the variability of TG 203 data, an existing TG 203 database was used, which was engaged earlier for re-assessing the correlation between TG 203 and TG 236 data (Sobanska et al., 2018). However, the TG 203 data in that database had not yet been filtered with the same quality criteria as the TG 236 data. A subsequent assessment of TG 203 variability used more extensive study quality filters and resulted in a reduction of a data set from originally 2936 studies on 1842 chemicals to 364 studies on 266 chemicals (Braunbeck et al., 2020). This reduction was of a similar magnitude to that seen in the TG 236 database (from 2065 to 123 chemicals, (Sobanska et al., 2018)) and indicates the same limitations for TG 203 as for TG 236 data in terms of data availability for lipophilic, volatile, reactive, inorganic and high molecular weight chemicals. Thus, in case the TG 236 database is considered too limited for reliably assessing the validity of TG 236, the same appears to hold true for TG 203.
9.2	For alternative methods, validation is obligatory for their acceptance and also available for the OECD FET TG 236 (Busquet et al., 2014; OECD, 2011, 2012b; Belanger et al., 2013) and the fish cell line test (Tanneberger et al., 2013; Fischer et al., 2019; Natsch et al., 2018; ISO, 2019). These validation studies support reliability in terms of intra- and inter-laboratory reproducibility of test results (see 10.2) and also relevance in terms of correlation with AFT data with slopes near 1 and intercepts close to 0. Some chemicals with a neurotoxic MoA and one chemical for which toxicity is increased due to bioactivation via alcohol dehydrogenase, appear to be less toxic in the two alternative methods compared to AFT. They represent outliers from the strong correlation observed between data from alternative methods and AFT. Such outliers may still be within the range of AFT variability, *inter alia*, due to the TG inherent species variability (see 10). However, with the aim of improving the current level of environmental protection, the development of an IATA, composed of several methods and information sources is currently envisaged at OECD level (WNT Project 2.54). Compared to *in vivo* methods, validation for reliability is relatively easy due to usually lower costs, higher throughput and 3Rs benefit (reducing the constraints for conducting a higher number of tests). Validation of relevance for real aquatic environments and with respect to interference with environmental factors is similarly difficult for alternative methods; however, computational approaches (as one kind of alternative method) can start to address this (see 6.2).
RELIABILITY (WITHIN AND BETWEEN LABORATORIES)	

10	The variability of protocol variants associated with acute fish toxicity testing was investigated and it was demonstrated that when including all test species, all life stages and all test conditions and not filtering for any type of exposure confirmation, the range between minimum and maximum 96 h LC_{50} values may be as great as six logarithmic units within the 44 substances analyzed. Even when data for rainbow trout only were considered, the range between the minimum and maximum 96 h LC_{50} value could still reach three logarithmic units (Hrovat et al., 2009).

Another data set and analysis (Belanger et al., 2013; Lammer et al., 2009) illustrates that the span of toxicity reported in the literature may be as large as 4 logarithmic units.

Two older ring trials indicate a maximum range of LC_{50} inter-laboratory variability of one logarithmic unit, if fish interspecies variability is excluded, but other variability such as fish size and exposure conditions (flow-through or static) is included. The two assessments were based on one chemical in two replicates and two fish species within 6 laboratories (Lemke, 1981) and another chemical in one or two replicates and one fish species within 13 laboratories (US EPA, 2001).

Below, in Table S1/10.0., a representative compilation of aquatic toxicity tests was devised to provide additional perspective on expected reproducibility from both accepted standardized assays used for regulatory purposes (macrophyte, algae, daphnid, and fish) and new (fish embryo, gill cell line) assays that have undergone international validation. The types of compounds used in ring trial and assay validation programs range from highly variable wastewater effluents, inorganic compounds, industrial organic compounds, pesticides, polymers, biocides and pharmaceuticals suggesting that the primary source of variability for the assessment of acute aquatic toxicity is the difference between the biological components used within the different test systems. Some ring trails included determination of assay conditions (e.g., Lemke (1981) who compared flow-through and static designs for daphnids and two fish species).

Tab. S1/10.0: Overview of representative levels on intra-laboratory (I) and inter-laboratory (E) repeatability acute toxicity tests for various taxa and test types
All results are expressed as Coefficients of Variation (%). Alternatives to acute fish toxicity are shaded in gray.

Species	Type of Assay	Guidance	Stressor	Coefficient of Variation (%)	Comments	Citation
Myriophyllum aquaticum (Aquatic macrophyte)	Sediment contact test	(ISO, 2013)	Natural and artificial sediments	E – 20.6-39.9	21 laboratories	Feiler et al., 2014
Myriophyllum spicatum (Aquatic macrophyte)	Aqueous growth inhibition	OECD Method validation	3,5 Dichlorophenol;2,4-Dichlorophenoxyacetic acid; Isoproturon	I – 10-30 E – 19.5-127.5	12 laboratories; 3 substances	OECD, 2014
Skeletonema costatum (marine diatom algae)	Acute inhibition	OECD 201	Potassium iodide	I – 7.9 E – 27.6	13 laboratories	Hantsveit, 1991
Daphnia magna (invertebrate)	Acute toxicity	ASTM E729-80	7 compounds	E – 21.4 (range of 2.5-77.4)	3 laboratories	Gersich et al., 1986
Daphnia magna (invertebrate)	Acute toxicity	USEPA 1002	Wastewater effluent	I – 18.1 (range of 4.2-8.9) E – 30.1	9 laboratories	Peltier, 1978
Daphnia magna (invertebrate)	Acute toxicity	ASTM E729-78	Silver nitrate, Endosulfan	E – 34.0-143.4	6 laboratories	Lemke, 1981
Pimephales promelas (Fathead minnow, fish)	Acute toxicity	ASTM E729-78	Endosulfan	E – 49.4	6 laboratories, flow through, 2 tests/ chemical	Lemke, 1981
Pimephales promelas (Fathead minnow, fish)	Acute toxicity	ASTM E729-78	Endosulfan	E – 35.1	6 laboratories, static 2 tests/ chemical	Lemke, 1981
Pimephales promelas (Fathead minnow, fish)	Acute toxicity	USEPA 1003	Potassium chloride	I – 7.6 E – 19.7	13 laboratories 1-2 tests/ chemical	US-EPA, 2001
Oncorhynchus mykiss (Rainbow trout, fish)	Acute toxicity	ASTM E729-78	Endosulfan	E – 49.7	6 laboratories Flow through 2 tests/ chemical	Lemke, 1981
Oncorhynchus mykiss	Acute toxicity	ASTM E729-78	Endosulfan	E – 43.5	6 laboratories 2 tests/	Lemke, 1981

						chemical	
(Rainbow trout, fish)							
Danio rerio (Zebrafish, fish)	Acute toxicity	OECD TG 236	20 compounds	I – 13.8 (range of 0-37.9) E – 20.6 (1.8-56.3)	10 laboratories 3 tests/ chemical		Busquet et al., 2014
Gill cell line	Acute cytotoxicity (RTgill-W1)	(ISO, 2019)	6 compounds	I – 15.5 E – 30.8	6 laboratories 3 tests/ chemical		Fischer et al., 2019

Without exception, inter-laboratory coefficients of variation exceed intra-laboratory coefficients of variation when both were available for comparisons within a ring trial.

From this summary, based on coefficient of variations and excluding the TG 203 inherent inter-species variability, it seems that both intra- and inter-laboratory repeatability are in a similar range for traditional ecotoxicity tests and alternative assays.

However, data sets referenced in the table above are larger for the alternative methods in terms of number of chemicals and replicates tested. Therefore, the slightly lower range estimates for variability in alternative methods compared to AFT, as derived from the references above, support the expectations for lower variability of alternative methods (see 10.2).

Moreover, a more comprehensive, and more recent analysis applying stringent data quality filters indicated that about 8% and 0.5% of 181 chemicals showed differences in AFT LC_{50} data by factors of > 10 or > 100, respectively, if interspecies variability is excluded. If the TG 203 inherent interspecies variability is included these percentages increase to about 15% and 10% of 178 chemicals (Braunbeck et al. 2020).

Also other work including AFT data for the neurotoxic biocide malathion indicates that AFT interspecies difference may be in the range of 4 orders of magnitude, depending on the chemical (Figure 5 in Fischer et al. 2019). For compounds that require activation, differences in the LC_{50} of > 50-fold have been identified for different species (Scholz et al., 2016).

An earlier analysis using less stringent AFT data-quality filters compared to Braunbeck at al. 2020 – used a database of 337 chemicals and 4 fish species. This study indicated an AFT interspecies variability between a factor of 1 and 43 for the mean, and a factor of between 8 and 95 for the 95% quantiles (Scholz et al., 2016; Sobanska et al., 2018).

10.1	Uncertainties:
	TG 203 is flexible regarding various species, and the test-species related water conditions (temperature, salinity, pH, water hardness), and test conditions (static, flow through, organism loading). This could result in variable results. Whether the use of stringent test parameters would result in a reduced variability is the subject of discussion and would require an appropriate systematic assessment of existing data. However, in the absence of a standard validation of TG 203, there is uncertainty regarding how the flexible elements within the test design may influence variability. Yet, variable LC_{50} values are accepted by OECD countries due to its Mutual Acceptance of Data (MAD) principle.
10.2	If the OECD guidance on good in vitro method practice (GIVIMP) is followed to develop robust protocols (OECD, 2018b), alternative methods may produce less variable and less uncertain results due to their higher standardization, improved basic study design (see 3.2, 4.2 and 5.2) and the availability of standard validation data. This increases data comparability, data reproducibility and reliability of classification, PNEC derivation and (PB)T criterion identification, which is an advantage for (international) regulation (see 16.1)
	In fact, it is noted that – compared to the AFT ring-trials with single species and one or two chemicals (US EPA, 2001; Lemke, 1981; first two columns in Tab. S1/10.2) – the data-sets for the validation of alternative methods are larger in terms of number of chemicals and replicates tested. Therefore, the slightly lower range estimates for variability in alternative methods compared to these first two AFT variability estimates (US EPA, 2001; Lemke, 1981) support the expectations for lower variability of alternative methods. The more comprehensive, most recent analysis indicates clearly highest range estimates (excluding and including interspecies variability), but this is also based on many more chemicals. It is noted that TG 203 recommends the use of different fish-species and this clearly further increases LC_{50} variability.

Tab. S1/10.2: Variability estimates for AFT and alternative methods based on LC_{50} range-estimates

	AFT – USEPA 1003 *excluding interspecies variability*	AFT – ASTM E729-78 *excluding interspecies variability*	AFT – TG 203 *excluding interspecies variability*	AFT – TG 203 *including interspecies variability*	ZFET – OECD TG 236	RTgill-W1 cell line test (ISO, 2019)
Dataset	1 chemical (potassium chloride) in 1-2 replicates, 13 laboratories, 1 fish species	1 chemical (endosulfan) in 2 replicates, 6 laboratories, 2 fish species	181 chemicals, 493 studies	53 chemicals, 178 studies	13 chemicals in 3 replicates, 4 laboratories	6 chemicals in 3 replicates, 5 laboratories
LC_{50} range between replicates and labs	max. factor = 3 for fathead minnow	max. factor = 13 for fathead minnow (3 from 24 studies > 10x) max. factor = 14 for rainbow	> factor 10 for 15 chemicals (8%) > factor 100 for 1 chemical (0.5%)	> factor 10 for 8 chemicals (15%) > factor 100 for 5 chemicals (10%)	~factor 4 for 2 chemicals < factor 3 for 11 chemicals	~factor 6 for 1 chemical < factor 3 for 5 chemicals

			trout (1 from 24 studies > 10x)				
	Reference	Table 9.14. in US EPA (2001)	Table 5 in Lemke (1981)	Braunbeck et al., 2020	Braunbeck et al., 2020	Table 16 in Annex V of OECD (2012b)	Supplement to Fischer et al., 2019

PREDICTIVE CAPACITY OR EXTRAPOLATION UNCERTAINTY

11	For a summary see 9.2 For semi-quantitative information see 2.1, last paragraph. For quantitative information see 6 and 6.1.
	TG 203 and alternative methods are conceptually similar with regard to the fact that data outputs from both types of methods need to be fed into extrapolation models that can provide information on potential effects and PNECs in the real environment.
	However alternatives allow improvements (see 6.2).

PROPRIETARY ASPECTS

12	none

REGULATORY USE

13	See 5 to 7.2, 16 and 16.1.

TECHNICAL LIMITATIONS

14	Using TG 203 for testing chemicals with slow bioaccumulation is highly questionable from an environmental relevance point of view. The acute exposure period may be too short for building up a critical body burden and hence the toxicity could be underestimated.
	It may also be discussed whether the current approach of testing for acute and/or chronic fish toxicity depending on yearly production volumes, exposure scenarios and results in acute fish tests is scientifically adequate at all. Previously it was suggested that testing with chemical specific exposure periods, in order to allow testing at conditions of steady state body burdens, could be scientifically more defensible (Sprague, 1969).[9]
14.1	*In vitro* methods and computational models are available that may inform on the potential for bioaccumulation in fish (OECD, 2018a), OECD QSAR toolbox, US-EPA TEST, VEGA, and potentially useful follow up assessment strategies.
15	Low water solubility, high volatility, chemical instability, mixtures and UVCBs.
	The OECD Guidance Document 23 for aqueous-phase aquatic toxicity testing provides recommendations for the testing of difficult-to-test test chemicals (OECD, 2019).
15.1	Computational modelling may allow extrapolation to a chemical space that would be technically impossible or difficult to test (Thomas et al., 2019).
	For alternative experimental methods, the same limitations with regard to testing difficult substances would apply. As for the AFT, appropriate approaches have been suggested to test difficult compounds such as, e.g., sealing culture dishes to avoid evaporation, frequent renewal of exposure solutions and/or saturation of the culture dishes by pre-incubation with chemicals, which is potentially relevant for highly hydrophobic substances (OECD, 2018b).

LIMITATIONS WITH REGARD TO REGULATORY APPLICABILITY

16	TG 203 is a vertebrate animal test based on lethality as an endpoint using a minimum of 42 animals beyond the embryonic stage per test in a full concentration-response study. These two aspects are ethically and politically problematic. The United States of America, the Netherlands and the United Kingdom have published high level policy plans for phasing out animal testing within the next 5 to 15 years[13]. In principle, the European Directive for the Protection of Animals used for Scientific Purposes aims to fully replace the use of animals in scientific procedures (Recital 10, 2010/63/EU). The European ban for animal testing for cosmetics is causing a precedence for this field for a growing number of other global regions[14].
	TG 203 can only be conducted in a low-throughput manner, since it requires rearing of fish to the suitable juvenile size, which may take several weeks, depending on the species. Testing as such requires 96 hours, not including the additional time for planning, preparation, assessment and reporting. The throughput is further limited by the large volumes and vessels required to conduct the test.

---

[13] US: A Strategic Roadmap for Establishing New Approaches to Evaluate the Safety of Chemicals and Medical Products in the United States. National Toxicology Program, U.S. Department of Health and Human Services.
https://ntp.niehs.nih.gov/whatwestudy/niceatm/natl-strategy/index.html (updated 07.05.2019; accessed 23.05.2020).
UK: Innovate UK. A non-animal technologies roadmap for the UK Advancing predictive biology.
https://nc3rs.org.uk/sites/default/files/documents/NonAnimalTechCO082_RYE_4_nrfinal2.pdf (published 10.11.2015; accessed 23.05.2020).
NL: Netherlands National Committee for the protection of animals used for scientific purposes. Transition to non-animal research on opportunities for the phasing out of animal procedures and the stimulation of innovation without laboratory animals. Published December, 12, 2016. file:///C:/Users/Gilly/Downloads/NCad+Opinion+Transition+to+non-animal+research.pdf (accessed 23.05.2020).
[14] Though this testing-ban has currently only manifested for human toxicity, the strong international political interest to end animal testing in the field of cosmetics has become obvious: Ban on animal testing. European Commission Internal Market, Industry, Entrepreneurship and SMEs. https://ec.europa.eu/growth/sectors/cosmetics/animal-testing_en (accessed 23.05.2020).

	The European strategy for a non-toxic environment (European Commission, 2017) states it would be desirable to test a larger proportion of the 100.000 chemicals currently on the market, which are still lacking reliable toxicity data[15]. Low assay throughput as a result of the time taken for the study and the large volumes of chemical and tank volumes represent a limitation to this goal. Moreover, there is regulatory interest for newer areas, like toxicity of nanomaterials[16], mixtures (European Commission, 2018) and bio-analytics in environmental aquatic media that may require acute aquatic toxicity estimates (Norberg-King et al., 2018; Schroeder et al., 2016). The design of "green chemistry" molecules may also require additional testing (Maertens et al., 2014).
	LC_{50} data variability (see 10 and 10.1) may translate (in up to 35% of cases, see 2.1) to variable PNECs and GHS classifications, for toxicologically similar chemicals, which is a disadvantage for global regulation.
16.1	Providing absolute environmentally relevant toxicity information for the extensive real-world environmental variability is extremely difficult and practically impossible to achieve with any method. From a scientific point of view, hazard identification (i.e., GHS classification and (PB)T criterion identification) and hazard characterization (i.e., concentration-response modeling and PNEC derivation) may support the comparison of the toxicity of chemicals. This will support a global reduction in environmental exposure to the most dangerous chemicals. From this perspective, alternative methods may provide the following advantages: ○ 3Rs benefits • Replacement of AFT in most cases by using a threshold approach-based IATA (including physico-chemical-data, algae and daphnids tests, fish-cell-tests, computational methods and other approaches), • Refinement by using FET instead of AFT in most cases within the threshold approach-based IATA, • Reduction in use of AFT and FET, by using a threshold approach-based IATA, (which also supports the use of limit tests for AFT and FET), ○ reducing costs (by using, for example, computational models, fish cell line test), ○ increasing the testing and assessment throughput (by using, for example, computational models, fish cell line test), ○ improved similarity of test designs and assessments (engaging IATAs, if necessary), ○ reduced variability and uncertainty thereof (see 3.2, 4.3, 5.2, 8.2, 9.2 and 10.2). This shall permit more reliable data generation in a more contemporary time frame for many more chemicals. Altogether, this may increase the comparability of test results between chemicals and thus global reliability of GHS classification, PNEC derivation and risk assessment. Such improved data quality may also better support the development of computational methods, which, if sufficiently valid, may be highly efficient for the ecotoxicity assessment of chemicals.

References

Arini, A., Mittal, K., Dornbos, P. et al. (2017). A cell-free testing platform to screen chemicals of potential neurotoxic concern across twenty vertebrate species. *Environ Toxicol Chem 36*, 3081-3090. doi:10.1002/etc.3880

Awkerman, J. A., Raimondo, S., Jackson, C. R. et al. (2014). Augmenting aquatic species sensitivity distributions with interspecies toxicity estimation models. *Environ Toxicol Chem 33*, 688-695. doi:10.1002/etc.2456

Bauer, J. F., Thomas, P. C., Fouchard, S. Y. et al. (2018a). A new classification algorithm based on mechanisms of action. *Comput Toxicol 5*, 8-15. doi:10.1016/j.comtox.2017.11.001

Bauer, J. F., Thomas, P. C., Fouchard, S. Y. et al. (2018b). High-accuracy prediction of mechanisms of action using structural alerts. *Comput Toxicol 7*, 36-45. doi:10.1016/j.comtox.2018.06.004

Bejarano, A. C., Raimondo, S. and Barron, M. G. (2017). Framework for optimizing selection of interspecies correlation estimation models to address species diversity and toxicity gaps in an aquatic database. *Environ Sci Technol 51*, 8158-8165. doi:10.1021/acs.est.7b01493

Belanger, S. E., Rawlings, J. M. and Carr, G. J. (2013). Use of fish embryo toxicity tests for the prediction of acute fish toxicity to chemicals. *Environ Toxicol Chem 32*, 1768-1783. doi:10.1002/etc.2244

Belanger, S. E., Sanderson, H., Embry, M. R. et al. (2015). It is time to develop ecological thresholds of toxicological concern to assist environmental hazard assessment. *Environ Toxicol Chem 34*, 2864-2869. doi:10.1002/etc.3132

Benfenati, E., Manganaro, A. and Gini, G. (2013). VEGA-QSAR: AI inside a platform for predictive toxicology. CEUR Workshop Proceedings. http://ceur-ws.org/Vol-1107/

Braunbeck, T., Böhler, S., Knörr, S. et al. (2020). Development of an OECD guidance document for the application of OECD Test Guideline 236 (Acute Fish Embryo Toxicity Test): The chorion structure and biotransformation capacities of zebrafish as boundary conditions for OECD Test Guideline 236 – German contributions to OECD project 2.54: Guidance Document on IATA for Fish Acute Toxicity Testing. UBA Texte 94/2020, in press. https://bit.ly/345BHsc

Busquet, F., Strecker, R., Rawlings, J. M. et al. (2014). OECD validation study to assess intra- and inter-laboratory reproducibility of the zebrafish embryo toxicity test for acute aquatic toxicity testing. *Regul Toxicol Pharmacol 69*, 496-511. doi:10.1016/j.yrtph.2014.05.018

---

[15]The estimate is based on the number of chemicals in the ECHA Classification and Labelling inventory, i.e., about 142,000 that should be on the European market. About 22,000 chemicals are registered for REACH in more than 1 ton per year, for which acute toxicity studies for daphnids and algae are required, but only substances registered above 10 tons per year have requirements for acute fish toxicity, these are about 7000 chemicals. Slightly more than 60% of the data-requirements were filled by experimental studies and the rest with read across, QSAR or Weight of Evidence assessments ECHA (2017a).

[16] Nanomaterials. European Commission Environment website. https://ec.europa.eu/environment/chemicals/nanotech/index_en.htm (updated 09.08.2019; accessed 23.05.2020).

Carr, G. J., Bailer, A. J., Rawlings, J. M. et al. (2018). On the impact of sample size on median lethal concentration estimation in acute fish toxicity testing: Is n = 7/group enough? *Environ Toxicol Chem 37*, 1565-1578. doi:10.1002/etc.4098

de-Vlaming, V. and Norberg-King, T. (1999). A review of single species toxicity tests: Are the tests reliable predictors of aquatic ecosystem community responses? EPA/600/R-97/114. https://bit.ly/3c1FI3T

Dimitrov, S., Detroyer, A., Piroird, C. et al. (2016). Accounting for data variability, a key factor in in vivo/in vitro relationships: application to the skin sensitization potency (in vivo LLNA versus in vitro DPRA) example. *J Appl Toxicol 36*, 1568-1578. doi:10.1002/jat.3318

ECHA (2008). Guidance on Information Requirements and Chemical Safety Assessment. Chapter R.10: Characterisation of dose [concentration]-response for environment. Guidance for the implementation of REACH. doi:10.4337/9781784711313.00012

ECHA (2017a). The use of alternatives to testing on animals for the REACH Regulation. Third report under Article 117(3) of the REACH Regulation. https://echa.europa.eu/documents/10162/13639/alternatives_test_animals_2017_en.pdf doi:10.2823/023078

ECHA (2017b). Guidance on Information Requirements and Chemical Safety Assessment. Chapter R.11: PBT/vPvB assessment. Version 3.0. https://echa.europa.eu/guidance-documents/guidance-on-information-requirements-and-chemical-safety-assessment

ECHA (2017c). Guidance on the Application of the CLP Criteria (Version 5.0). https://echa.europa.eu/guidance-documents/guidance-on-clp

European Commission (2017). Strategy for Non-toxic-environment, Executive summary. http://ec.europa.eu/environment/chemicals/non-toxic/index_en.htm

European Commission (2018). Something from nothing? Ensuring the safety of chemical mixtures. JRC Science for Policy Brief. doi:10.2760/618648

Feiler, U., Ratte, M., Arts, G. et al. (2014). Inter-laboratory trial of a standardized sediment contact test with the aquatic plant Myriophyllum aquaticum (ISO 16191). *Environ Toxicol Chem 33*, 662-670. doi:10.1002/etc.2483

Fischer, M., Belanger, S. E., Berckmans, P. et al. (2019). Repeatability and reproducibility of the RTgill-W1 cell line assay for predicting fish acute toxicity. *Toxicol Sci 169*, 353-364. doi:10.1093/toxsci/kfz057

Gersich, F. M., Blanchard, F. A., Applegath, S. L. et al. (1986). The precision of daphnid (Daphnia magna Straus, 1820) static acute toxicity tests. *Arch Environ Contam Toxicol 15*, 741-749. doi:10.1007/BF01054921

Hahn, T., Diamond, J., Dobson, S. et al. (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. doi:10.1002/ieam.1473

Hoekzema, C. C., Murk, A. J., Van de Waart, B. J. et al. (2006). Alternative approaches can greatly reduce the number of fish used for acute toxicity testing. *Environ Toxicol Chem 25*, 1322-1325. doi:10.1897/05-293R2.1

Hrovat, M., Segner, H. and Jeram, S. (2009). Variability of in vivo fish acute toxicity data. *Regul Toxicol Pharmacol 54*, 294-300. doi:10.1016/j.yrtph.2009.05.013

Hutchinson, T. H., Barrett, S., Buzby, M. et al. (2003). A strategy to reduce the numbers of fish used in acute ecotoxicity testing of pharmaceuticals. *Environ Toxicol Chem 22*, 3031-3036. doi:10.1897/02-558

ISO (2013). Water quality – Determination of the toxic effect of sediment on the growth behaviour of Myriophyllum aquaticum. ISO 16191:2013. https://www.iso.org/standard/55809.html

ISO (2019). Water quality – Determination of acute toxicity of water samples and chemicals to a fish gill cell line (RTgill-W1). ISO 21115:2019. https://www.iso.org/standard/69933.html

Jeram, S., Sintes, J. M., Halder, M. et al. (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*, 218-224. doi:10.1016/j.yrtph.2005.04.005

Kluver, N., Ortmann, J., Paschke, H. et al. (2014). Transient overexpression of adh8a increases allyl alcohol toxicity in zebrafish embryos. *PLoS One 9*, e90619. doi:10.1371/journal.pone.0090619

Kluver, N., Konig, M., Ortmann, J. et al. (2015). Fish embryo toxicity test: identification of compounds with weak toxicity and analysis of behavioral effects to improve prediction of acute toxicity for neurotoxic compounds. *Environ Sci Technol 49*, 7002-7011. doi:10.1021/acs.est.5b01910

Knillmann, S., Stampfli, N. C., Beketov, M. A. et al. (2012). Intraspecific competition increases toxicant effects in outdoor pond microcosms. *Ecotoxicology 21*, 1857-1866. doi:10.1007/s10646-012-0919-y

Knobel, M., Busser, F. J., Rico-Rico, A. et al. (2012). Predicting adult fish acute lethality with the zebrafish embryo: relevance of test duration, endpoints, compound properties, and exposure concentration analysis. *Environ Sci Technol 46*, 9690-9700. doi:10.1021/es301729q

Lammer, E., Carr, G. J., Wendler, K. et al. (2009). Is the fish embryo toxicity test (FET) with the zebrafish (Danio rerio) a potential alternative for the fish acute toxicity test? *Comp Biochem Physiol C Toxicol Pharmacol 149*, 196-209. doi:10.1016/j.cbpc.2008.11.006

Lemke, A. (1981). Interlaboratory Comparision: Acute Testing Set. US-EPA EPA-600/S3-81-005 https://nepis.epa.gov/Exe/ZyPURL.cgi?Dockey=2000TSC7.TXT

Lillicrap, A., Moe, S. J., Wolf, R. et al. (2020). Evaluation of a Bayesian network for predicting acute fish toxicity from fish embryo toxicity data. *Integr Environ Assess Manag 16*, 452-460. doi:10.1002/ieam.4258

Maertens, A., Anastas, N., Spencer, P. J. et al. (2014). Green toxicology. *ALTEX 31*, 243-249. doi:10.14573/altex.1406181

Moe, S. J., Madsen, A. L., Connors, K. A. et al. (2020). Development of a hybrid Bayesian network model for predicting acute fish toxicity using multiple lines of evidence. *Environ Modell Softw 126*, 104655. doi:10.1016/j.envsoft.2020.104655

Mora, C., Tittensor, D. P., Adl, S. et al. (2011). How many species are there on earth and in the ocean? *PLoS Biol 9*, e1001127. doi:10.1371/journal.pbio.1001127

Natsch, A., Laue, H., Haupt, T. et al. (2018). Accurate prediction of acute fish toxicity of fragrance chemicals with the RTgill-W1 cell assay. *Environ Toxicol Chem 37*, 931-941. doi:10.1002/etc.4027

Norberg-King, T. J., Embry, M. R., Belanger, S. E. et al. (2018). An international perspective on the tools and concepts for effluent toxicity assessments in the context of animal alternatives: Reduction in vertebrate use. *Environ Toxicol Chem 37*, 2745-2757. doi:10.1002/etc.4259

OECD (2002). Guidance Document on the Recognition, Assessment and Use of Clinical Signs as Humane Endpoints for Experimental Animals Used in Safety Evaluation. *Series of Testing and Assessment, No. 19*. OECD Publishing, Paris. doi:10.1787/9789264078376-en

OECD (2011). Validation Report (Phase 1) for the Zebrafish Embryo Toxicity Test. *Series of Testing and Assessment, No. 157*. OECD Publishing, Paris. http://www.oecd.org/env/ehs/testing/48572244.pdf

OECD (2012a). Fish Testing Framework. *Series of Testing and Assessment, No. 171*. OECD Publishing, Paris. doi:10.1787/9789264221437-en

OECD (2012b). Validation Report (Phase 2) for the Zebrafish Embryo Toxicity Test. *Series of Testing and Assessment, No. 179*. OECD Publishing, Paris. doi:10.1787/9789264221437-en

OECD (2014). Myriophyllum Spicatum Toxicity Test: Results of an Inter-Laboratory Ring Test Using a Sediment-Free Test System – Ring Test Report. *Series of Testing and Assessment, No. 205*. OECD Publishing, Paris. https://bit.ly/3nmX5Ro

OECD (2016). Guidance Document on the Reporting of Defined Approaches to be Used Within Integrated Approaches to Testing and Assessment. *Series on Testing and Assessment, No. 255*. OECD Publishing, Paris. doi:10.1787/9789264274822-en

OECD (2018a). Guidance Document on the Determination of In Vitro Intrinsic Clearance Using Cryopreserved Hepatocytes (RT-HEP) or Liver S9 Sub-Cellular Fractions (RT-S9) from Rainbow Trout and Extrapolation to In Vivo Intrinsic Clearance. *Series of Testing and Assessment, No. 280*. OECD Publishing, Paris. http://www.oecd.org/chemicalsafety/testing/series-testing-assessment-publications-number.htm

OECD (2018b). Guidance Document on Good In Vitro Method Practice (GIVIMP). *Series of Testing and Assessment, No. 286*. OECD Publishing, Paris. doi:10.1787/9789264304796-21-en

OECD (2019). Guidance Document on Aqueous-Phase Aquatic Toxicity Testing of Difficult Test Chemicals. *Series of Testing and Assessment, No. 23* (second edition). OECD Publishing, Paris. doi:10.1787/0ed2f88e-en

Peltier, W. (1978). Methods for Measuring the Acute Toxicity of Effluents to Aquatic Organisms. EPA 600/4-78-012. https://bit.ly/3ab15kn

Rawlings, J. M., Belanger, S. E., Connors, K. A. et al. (2019). Fish embryo tests and acute fish toxicity tests are interchangeable in the application of the threshold approach. *Environ Toxicol Chem 38*, 671-681. doi:10.1002/etc.4351

Rufli, H. (2012). Introduction of moribund category to OECD fish acute test and its effect on suffering and LC50 values. *Environ Toxicol Chem 31*, 1107-1112. doi:10.1002/etc.1779

Schlenk, D., Celander, M., Gallagher, E. et al. (2008). Biotransformation in fishes. In R. T. Di Giulio and D. E. Hinton (eds.), *The Toxicology of Fishes*. Boca Raton, USA: CRC Press. doi:10.1201/9780203647295.ch4

Scholz, S., Klüver, N. and Kühne, R. (2016). Analysis of the relevance and adequateness of using Fish Embryo Acute Toxicity (FET) Test Guidance (OECD 236) to fulfil the information requirements and addressing concerns under REACH. Report ECHA-UFZ contract ECHA/2014/341, 1-105.

Schroeder, A. L., Ankley, G. T., Houck, K. A. et al. (2016). Environmental surveillance and monitoring – The next frontiers for high-throughput toxicology. *Environ Toxicol Chem 35*, 513-525. doi:10.1002/etc.3309

Sobanska, M., Scholz, S., Nyman, A. M. et al. (2018). Applicability of the fish embryo acute toxicity (FET) test (OECD 236) in the regulatory context of registration, evaluation, authorisation, and restriction of chemicals (REACH). *Environ Toxicol Chem 37*, 657-670. doi:10.1002/etc.4055

Sprague, J. B. (1969). Measurement of pollutant toxicity to fish I. Bioassay methods for acute toxicity. *Water Res 3*, 793-821. doi:10.1016/0043-1354(69)90050-5

Stadnicka-Michalak, J., Weiss, F. T., Fischer, M. et al. (2018). Biotransformation of Benzo[a]pyrene by three rainbow trout (Onchorhynchus mykiss) cell lines and extrapolation to derive a fish bioconcentration factor. *Environ Sci Technol 52*, 3091-3100. doi:10.1021/acs.est.7b04548

Stengel, D., Wahby, S. and Braunbeck, T. (2018). In search of a comprehensible set of endpoints for the routine monitoring of neurotoxicity in vertebrates: Sensory perception and nerve transmission in zebrafish (Danio rerio) embryos. *Environ Sci Pollut Res Int 25*, 4066-4084. doi:10.1007/s11356-017-0399-y

Tanneberger, K., Knobel, M., Busser, F. J. et al. (2013). Predicting fish acute toxicity using a fish gill cell line-based toxicity assay. *Environ Sci Technol 47*, 1110-1119. doi:10.1021/es303505z

Thomas, P. C., Bicherel, P. and Bauer, F. J. (2019). How in silico and QSAR approaches can increase confidence in environmental hazard and risk assessment. *Integr Environ Assess Manag 15*, 40-50. doi:10.1002/ieam.4108

Tierbach, A., Groh, K. J., Schonenberger, R. et al. (2018). Glutathione S-transferase protein expression in different life stages of zebrafish (Danio rerio). *Toxicol Sci 162*, 702-712. doi:10.1093/toxsci/kfx293

US-EPA (2001). Final Report: Interlaboratory Variability Study of EPA Short-term Chronic and Acute Whole Effluent Toxicity Test Methods, Vol. 1. EPA 821-B-01-004. Office of Water, U.S. Environmental Protection Agency, Washington, D.C https://bit.ly/3nraCaU

Volz, D. C., Belanger, S., Embry, M. et al. (2011). Adverse outcome pathways during early fish development: A conceptual framework for identification of chemical screening and prioritization strategies. *Toxicol Sci 123*, 349-358. doi:10.1093/toxsci/kfr185

Zhao, Q., De Laender, F. and Van Den Brink, P. J. (2020). Community composition modifies direct and indirect effects of pesticides in freshwater food webs. *Sci Total Environ 739*, 139531. doi:10.1016/j.scitotenv.2020.139531

Zindler, F., Beedgen, F., Brandt, D. et al. (2019). Analysis of tail coiling activity of zebrafish (Danio rerio) embryos allows for the differentiation of neurotoxicants with different modes of action. *Ecotoxicol Environ Saf 186*, 109754. doi:10.1016/j.ecoenv.2019.109754