Weight of evidence tools in the prediction of acute fish toxicity

Scott E. Belanger,Adam D. Lillicrap,S. Jannicke Moe,Raoul Wolf,Kristin Connors,Michelle R. Embry
First published: 20 January 2022 https://doi.org/10.1002/ieam.4581

Abstract

Acute fish toxicity (AFT) is a key endpoint in nearly all regulatory implementations of environmental hazard assessments of chemicals globally. While it is an early tier assay, the AFT assay is complex and utilizes many juvenile fish each year for the registration and assessment of chemicals. Thus, it is imperative to seek animal alternative approaches to replace or reduce animal use for environmental hazard assessments. A Bayesian Network (BN) model has been developed that brings together a suite of lines of evidence (LoEs) to produce a probabilistic estimate of acute fish toxicity without the testing of additional juvenile fish. Lines of evidence include chemical descriptors, mode of action (MOA) assignment, knowledge of algal and daphnid acute toxicity, and animal alternative assays such as fish embryo tests and *in vitro* fish assays (e.g., gill cytotoxicity). The effort also includes retrieval, assessment and curation of quality acute fish toxicity data as these act as the baseline of comparison to model outputs. An ideal outcome of this effort would be to have global applicability, acceptance and uptake, relevance for predominant fish species used in chemical assessments, be expandable to allow incorporation of future knowledge and data be publicly available. The BN model can be conceived as having incorporated principles of tiered assessment and whose outcomes will be directed by the available evidence in combination with prior information. We demonstrate that as additional evidence is included into prediction of a given chemical's ecotoxicity profile, both the accuracy and the precision of the predicted AFT can increase. Ultimately an improved environmental hazard assessment will be achieved.

Introduction

Weight of Evidence (WoE) is frequently cited as being necessary for a wide variety of decision-making needs due to the complexity of environmental data (Hall et al. 2017). Understanding environmental fate, hazard and exposure for any well studied chemical will normally reveal the presence of conflicting data and the presence of novel test systems and endpoints. A WoE analysis of such diverse data requires judgment by technical experts with a range of expertise and philosophical leanings. While standardization of testing procedures can help somewhat in this regard, as the knowledge of individual chemicals' properties expands or uses of a chemical changes, it is appropriate to tailor the assessment of the chemical to those aspects which possess the greatest likelihood of informing decision-making and not be overly narrow in the application of standard test procedures. New endpoints or understanding subtle aspects of how a chemical interacts with biota will inject novel information into the decision-making process. Well known examples of this phenomenon are the use and application of genomic techniques in human and environmental assessments and the emergence of endocrine disruption as an environmental concern. New information may cause scientists to reassess what they thought they already knew or provide new avenues to consider when assessing a chemical entity.

Another relatively new and major area (for environmental assessment) is the use of animal alternative assays, often referred to as New Approach Methodologies or NAMs, to address endpoints of regulatory concern. Lillicrap et al. (2016) reviewed the general state-of the-science for various endpoints that require the use of fish for environmental assessment including bioaccumulation, acute and chronic toxicity and endocrine disruption. The acute fish toxicity (AFT) test (OECD 203, OECD 2019) is the most frequently used assay because it is a requirement in nearly all global regulatory schemes for the purposes of risk assessment and classification and labelling of chemicals (Burden et al. 2019). AFT tests account for the most fish used to satisfy regulatory testing requirements. While chronic (long-term) fish toxicity and bioaccumulation assays use more organisms per test, they are also used at a much lower frequency than AFT in the assessment of chemicals. Acute fish toxicity for effluent assessments in many countries utilize a large number of organisms and the employment of fish in research endeavors is a larger source of fish use as well (ECHA 2017). In our estimation, effluent testing

and research likely account for greater numbers than that where acute fish toxicity is a requirement for chemical registration (Norberg-King et al. 2018; ECHA 2017).

Bioethical concerns are a global driver for reducing the number of fish used in acute toxicity testing and replacing these tests with alternative *in vivo,* as well as *in vitro* and *in silico* approaches is a key objective. A wide range of regulatory, legal, and bioethical variability across the world makes the replacement of fish testing with NAMs extremely difficult. However, pressure on registrants of chemicals in commerce also continues to mount. In the absence of globally accepted recognition of replacement, reduction and refinement efforts, a reality is that acute fish toxicity tests may be required for registering a chemical in one region but not in another. Such dichotomies could be eliminated if evidence can be provided to convince scientists, and in particular regulatory authorities, that direct testing using juvenile fish (OECD 203) is unnecessary to decide if fish are the most sensitive taxonomic group for determining hazard . Lammer et al. (2009) and Belanger et al. (2013) provided substantial evidence that toxicity of chemicals to fish embryos was equivalent to typical fish acute toxicity. This assertion was challenged by Scholz et al. (2016), resulting in a subsequent opinion rendered by the ECHA (European Chemicals Agency) (2017) to industry and interested scientists to produce a WoE approach to fully support the use of fish embryo toxicity data (and other alternative assays/approaches) in place of AFT data in European chemical registrations.

Review of on-going work regarding AFT

Efforts to address the execution, utility, and interpretation of alternatives to acute fish toxicity have been on-going for almost 30 years. Early efforts by Nagel (1994) and colleagues led to a formalized fish embryo test for effluent testing which was standardized as an ISO Guideline in the mid 2000's and subsequently adopted as an OECD Test Guideline in 2013 (OECD 236; OECD 2013). The fish embryo test relies on certain apical endpoints to accurately predict lethality including coagulation (death), lack of heartbeat, lack of somite development, and non-detachment of the tail. Lammer et al. (2009) and Belanger et al. (2013) compared the fish embryo toxicity (FET) to the AFT and found nearly a 1:1 concordance for a large number of chemicals across a wide range of modes of action (Figure 1). This work formed the basis of the assertion by some European chemical registrants to waive the AFT in lieu of the FET, for which

ECHA since clarified could be done but in a broader WoE context. As a complicating matter, recent simulation studies of FET versus AFT study designs clearly show that the OECD recommended minimum fish per concentration (n=7) compromises certainty in the LC50 and confidence limit estimates relative to the study design recommended for fish embryos (n=20 per concentration with positive and negative controls) (Carr et al. 2018). Recent efforts to improve the approach to AFT in the OECD 203 Test Guideline process includes utilization of clinical signs that predict mortality (i.e., moribundity) and shortening the test duration (Katsiadaki et al. 2022).

Another alternative method is the implementation of the Threshold Approach (OECD 126, explained more fully below). The approach utilizes daphnid and algae toxicity tests to select the concentration at which to expose fish in a limit (threshold) test. Such an approach has been suggested as a means to reduce the use of fish by as much as 70% (Hutchinson et al. 2003). Although not explicitly expressed by Hutchinson et al. (2003), the Threshold Approach, can be considered a small WoE approach as it draws upon a variety of ecotoxicological insights to develop hazard conclusions.

Paparella et al. (2021) reviewed the known state of uncertainty and limitations of the acute fish toxicity test. The authors rightfully contend that the AFT has never been validated in the same sense that current OECD Test Guidelines require and as is outlined in OECD validation program documents under the Test Guidelines Programme (OECD 2005). The AFT, as currently practiced, has considerable drawbacks. Large variabilities in toxicity data for the same compound and species across laboratories are well documented in the literature (Belanger et al. 2013) and yet, AFT results are normally viewed as the gold standard against which alternative assays are judged. Regulatory inertia, tradition, and a lack of familiarity with strengths and weaknesses provided by alternative assays all seem to play a significant role in the lack of regulatory acceptance to move beyond *in vivo* testing (Lillicrap et al. 2016). Many advantages are afforded by alternative assays including increased statistical power, improved interlaboratory calibration and validation for new, more robust assays, higher throughput, and improved mechanistic insights to the chemical's activity (Carr et al. 2019; Lillicrap et al. 2016, Paparella et al. 2021). Combining these additional tools in a WoE approach has the potential to significantly improve the environmental hazard and risk assessment of chemicals rather than relying on only

whether an organism simply lives or dies, which at the population level is too crude to ensure the protection of the environment. While acute toxicity may be useful for generating information on potency of various chemicals, it has lower relevance for environmental hazard assessment where longer term, lower-level exposures and prediction of long-term effects and environmental behaviors are more important. Yet, the AFT assay remains a mandatory requirement in most regulatory schemes.

AFT is but one of several *in vivo* fish assays that are used to inform environmental decisions of chemicals. QSARs, chronic (long-term) fish toxicity, bioaccumulation, endocrine disruption and field surveys all find a home in the comprehensive assessment of potential environmental perturbations and chemical exposures to fish. In 2012, the OECD brought together a group of scientists to consider the integration of all assays utilizing fish, with the goal of more efficiently using those assays to inform each other and reduce testing burdens resulting in the Fish Framework document (OECD 212). Fish embryo testing was foreseen as a future input into the fish assay toolkit (and was also identified as such as early as 2008 in guidance documents for REACH; ECHA 2008). Figure 2 displays one possible outcome of an integrated view on fish testing. As with the discussions above, this is also a form of WoE which is mentioned prominently throughout the Fish Framework document and its recommendations for future action. The OECD workshop identified that the following strategies could be implemented to reduce *in vivo* testing: limit tests, Threshold approach, step-down approach, screening methodologies that do not utilize animals, (such as (Q)SAR tools, *in vitro* assays, or read-across), and Fish embryo tests (FET). The FET is considered a replacement or refinement assay depending on the regulatory jurisdiction owing to the utilization of the non-exogenous feeding, embryonic stage of development (Strahle et al. 2012).

Clearly an integrative process is needed to holistically address the types of information and data available to inform the acute toxicity of chemicals to fish. The sources of information are diverse, vary in quantitative rigor, and chemical coverage (when considering for example, different types of alternative assays). In acknowledgement of this, Moe et al. (2020) and Lillicrap et al. (2020) initiated an effort to construct a quantitative model to integrate all sources of information using a Bayesian network approach. A preliminary version of the model is being evaluated as a proof of concept to inform development of a suitable structure that can be

amplified with increasing sophistication of modelling, chemical coverage, endpoints, and types of information that could be useful in the future. The goal is to provide a regulatory actionable decision-support system for WoE in acute fish toxicity. Formalized as a project under the CEFIC LRI Programme called SWiFT (***S***trengthening ***W***eight of Ev***i***dence for Acute ***F***ish ***T***oxicity https://www.niva.no/swift), the Bayesian network model is a combination of tiered and probabilistic approaches. SWiFT intends to: 1) build a fully curated acute fish toxicity (AFT) database; 2) develop new (LoEs) for a Bayesian network (BN) model to predict acute fish toxicity; 3) develop and evaluate the BN model as a WoE framework to support AFT replacement; 4) develop a web interface for public access to the BN model; and, 5) produce guidance how to use the BN as a WoE tool.

Conceptual description of AFT lines of evidence

Predicting acute fish toxicity can draw from numerous LoEs and information. Typically, a complete understanding of a compound's physical and chemical properties is considered essential. Compound purity, solubility, molecular weight, sorptivity, pKa and other factors are ideally known and measured prior to testing (Schirmer et al. 2006). In addition, means to quantify in-test exposures under various water chemistry conditions is important as it is well known that exposure verification in *in vitro* and *in vivo* tests sheds important light on the comparability of measures between tests and across different compounds domains. That said, it is still uncommon to employ exposure verification, especially for difficult-to-test substances, and it is rare to summarize data for a compound where all possibly informative assays have equivalent levels of exposure verification (Belanger et al. 2013; Sobanska et al. 2018). This aspect alone has led researchers to employ varying levels of critical assessment of historical and current data when devising comparisons of alternative assays to the "gold standard" AFT (OECD 236) (Paparella et al. 2021). In this aspect, one can see how WoE processes could be advantageous to sort out the most relevant and strongest central tendencies.

As mentioned previously, the FET test has been a major contributor for providing data to predict the AFT using an alternative approach (OECD 2013; Belanger et al. 2013). Similarly, an even more recent effort to establish the utility of a fish gill cell line to evaluate cytotoxicity of chemicals, and by inference predict acute fish toxicity, has been developed (ISO 2019; Fischer et

175 al. 2019; OECD 249, OECD 2021). Thus, three closely related assays (acute fish toxicity, fish
176 gill cytotoxicity, and fish embryo toxicity) may be cross-referenced to inform acute fish toxicity
177 of previously untested chemicals. Belanger et al. (2013), Tannenberger et al. (2013), Sobanska et
178 al. (2018), and Fischer et al. (2019) each explored different aspects of the complexities involved
179 through comparisons of outcomes amongst these assays. Belanger et al. (Figure 1) compared the
180 toxicity of approximately 150 different compounds to FET and AFT (tests on the same
181 compound) and found a near 1:1 concordance. However, variation in the effect data for the same
182 chemical sometimes spanned orders of magnitude within either the AFT (more frequently)
183 versus the FET. This is likely due to the heterogeneity particularly in the AFT data with respect
184 to species choice (many species are recommended for use by OECD), test conditions, organism
185 size, water source, and utilization of analytical (or not) confirmation. The comparisons were
186 performed using orthogonal regression, a multivariate approach that allows for variation in both
187 x and y dimensions (unlike standard linear regression where the independent x-variable has fixed
188 values and the dependent y-variable is predicted with variation). Tannenberger et al. (2013)
189 developed a different database where the AFT focused solely on fathead minnow, which
190 constrains inter-species variability but simultaneously constrains AFT comparisons using species
191 fully recommended and endorsed by OECD (OECD 2019). Figure 3A shows comparisons for
192 cytotoxicity and *in vivo* toxicity for 37 compounds. Some, but not all, toxicity data in this Figure
193 are also presented in that of Figure 1. Fischer et al. (2019), as part of the international round
194 robin trial for the gill cytotoxicity assay, also summarized the comparative *in vitro* gill toxicity to
195 the AFT (Fig. 3B). Again, as the gill cytotoxicity assay is relatively new, the breadth of coverage
196 is less (albeit rapidly growing, cf. Natsch et al. 2018) and will partially intersect with that of the
197 FET and AFT. Importantly, the FET and gill cytotoxicity assay have both undergone
198 international validation and extensive ring trials to quantify their intra- and inter-laboratory
199 variabilities as well as a demonstration of transferability to naïve laboratories and is now
200 presently identified as OECD Test Guideline 249, “Fish Cell Line Acute Toxicity: The RTgill-
201 W1 cell line assay”. (Busquet et al. 2014; Fischer et al. 2019; OECD 2021). Sobanska and
202 colleagues presented a different view of the FET-AFT relationship that highlights some
203 uncertainties in the newer assay (see Fig 4 and Sobanska et al. 2018). Quantification of chemical
204 exposures identified a number of uncertainties which were subsequently resolved (Birke and
205 Scholz 2019), but others remain, especially those with specific modes of action (see discussion

206 also below). As debated at the FET Workshop hosted by ECHA in 2017 (ECHA 2017), a
207 mechanism is urgently needed to weigh results, predictions and ultimately derive predicted AFT
208 when an *in vivo* test is not available to meet increasing expectations by society, regulators and
209 scientists with regards to improved animal welfare.

210 An unrelated, but similarly complex module of a potential WoE scheme is that of the threshold
211 approach (TA, OECD 126). The Threshold Approach, by itself, has also been posited as a WoE
212 approach. In the Threshold Approach, acute toxicity tests are first performed on algae
213 (unspecified species, but likely one of the small number of internationally accepted test species
214 in OECD 201, OECD 2011) and *Daphnia* sp. (OECD 202, 2004). The more sensitive species of
215 algae and *Daphnia* is identified by the lower of the two effect values (the concentration causing a
216 50% effect - EC50). This concentration is then used as the test concentration in an acute toxicity
217 limit test (i.e., 1 concentration at the threshold value of toxicity). If fish are not affected at this
218 threshold concentration (i.e., 0% mortality), the assay confirms that fish are less sensitive and the
219 assessor can proceed with confidence that the lower hazard value from the algae and daphnid
220 tests will adequately inform the risk assessment. If fish are affected at the threshold
221 concentration, then a full acute fish toxicity test is required. An easy extension to this thinking is
222 to replace the AFT with that of the FET instead, thereby excluding the need for juvenile fish
223 altogether. Rawlings et al. (2019) explored various permutations of algae-daphnid-AFT-FET
224 testing to determine if the use of the FET or AFT impacted either the most sensitive hazard value
225 or GHS classification. The authors conclude that there is no distinction (advantage) of using the
226 AFT instead of the FET in the threshold approach and they can be considered equivalent. Such
227 comparisons can also be used to assist testing prioritization and weightings applied to each assay
228 for assessing overall acute fish toxicity. Rawlings and Belanger (personal communication,
229 unpublished data) extended the threshold approach concepts to QSARs and found that while
230 quantitative differences existed when using QSARs versus “real” data, the trends were
231 remarkably similar albeit slightly higher uncertainty.

232 Scientists researching alternative methods consistently agree that identification of the mode of
233 action (MOA) is an essential aspect of understanding toxicity relationships among chemicals and
234 assays (Kienzler et al. 2017). Schirmer et al. (2006), Belanger et al. (2013), Tannenberger et al.
235 (2013), Sobanska et al. (2018) and Fisher et al. (2019) use MOA assignments in parsing out

trends within the various databases used for comparing fish acute toxicity across compounds. This is particularly important as it is becoming well established that neurotoxicants represent a unique class of chemicals based on their MOA that are less well predicted by alternative assays (suggesting somewhat more frequently the FET and cytotoxicity tests, for example, are less sensitive to this particular mode of action) compared to the AFT. This is in conflict with the observation that neurotoxicants are also somewhat less toxic to fish overall than they are to invertebrates (Threshold Approach thinking, Connors et al. 2019). MOA is therefore critical in the development of WoE for predicting AFT. Kienzler et al. (2017) developed a MOA comparison for hundreds of compounds using several assignment schemes. These, like other databases, are built from different data sets, include different compounds, and were truly developed for different purposes or emphases. Subsequently, Kienzler et al. (2019) established a consensus approach to unify the MOA schemes with a single outcome based on physical-chemical information and multifaceted expert opinions, again forming a type of WoE.

Other factors may contribute to informing fish toxicity including collation of highly diverse fish metabolism and physiology information that can be useful to support MOA and understanding the potential for metabolic activation of unusual or specific compounds required to invoke toxicity to fish. While limited cases for such activation exist (allyl alcohol for example, see Kluver et al. 2014), these can explain outliers or trend busters for FET-AFT or cytotoxicity-AFT relationships (UBA 2020). Evolving neural toxicity assays, such as the quantification of the "touch-evoke" response or other behavioral endpoints using video tracking software, can more definitively determine the likelihood that a toxicant is or is not a neurotoxicant and thus lead to reliance on certain LoEs (e.g., the FET) versus others in predicting the AFT. The assignment of MOA will remain a significant area of research and will likely expand to include additional behavioral assessments such as swimming speed, endurance, and hyperactivity/hypoactivity.

The Acute Fish Toxicity Bayesian Network Model as a Path to Resolution

As indicated previously, Moe et al. (2020) and Lillicrap et al. (2020) provided the framework for developing a quantitative predictor of AFT using a Bayesian network (BN) model including multiple existing and proposed LoEs. BNs are graphical, probabilistic and potentially causal models, and have been used increasingly in hazard and risk assessment during the last decades

(Moe et al. 2021a). One of the strengths is the flexibility in model structure, which enables alignment of BN models with established frameworks such as adverse outcome pathways (Moe et al. 2021b) and the relative risk model (Landis 2021).

The BN model was developed using established WOE frameworks and guidance (e.g., Suter et al. 2017b, EFSA Scientific Committee et al. 2017) and is consistent with that described in this IEAM Special Series. The model gathers, weighs and integrates a wide range of LoEs that are used to develop a prediction of acute fish toxicity with a specified probability. A simplified graphical version of the preliminary model is shown in Figure 5. A database was built using all possible chemicals for which data on an alternative biological assay, relevant to AFT, are available. If only AFT and QSAR data were available for a compound, it is not considered in this database until an alternative assay is performed. The goal is to provide predictive support for acute fish toxicity using animal alternative toxicity tests. An expectation of the BN approach was that inclusion of data from more LoEs would result in a more accurate AFT prediction (that is, its central tendency being closer to the measured AFT). Model accuracy was defined by correctly predicting the toxicity interval of the AFT data using multiple LoEs. The accuracy rate of the BN model prediction was in the range of 69-80% when using available data for all LoEs as far as possible. A diverse array of chemicals and MOAs was used to test model predictions to assess if the model could be broadly applicable (Moe et al 2020). The strictest quality criterion resulted in a set of 20 chemicals which included various industrial compounds, surfactants, pesticides, and pharmaceuticals (see Table 2 of Lillicrap et al. 2020). For this subset, the BN predicted the correct toxicity interval for 80% of the chemicals evaluated. For the remaining 20% of the chemicals, daphnid or algae data were always more sensitive than fish, which means that daphnid or algae data would have driven any subsequent environmental hazard assessment or GHS classification. The evaluation also confirmed that the use of FET data to replace AFT data was justified in that GHS classification or toxicity interval predictions were the same.

Table 1 provides a summary of potential LoEs and their present likely strengths to support a WoE determination of AFT. Fish acute toxicity QSARs vary widely in the breadth of underlying data used in their development and the predictive chemical attribute also varies, but is most commonly based on log Kow. Through on-going efforts, such as those of EU JRC QSAR validation requirements for REACH (OECD 2004; EU JRC 2014), the aim is to provide a

295 baseline from which valid and accepted QSAR models can be derived and used for the BN. FET
296 and fish gill cytotoxicity are considered strong, direct evidence in support of AFT based on
297 earlier reviews of these assays. The Threshold Approach (TA) , as described earlier, is
298 considered supportive and indirect evidence of AFT. The outcomes of the TA is frequently
299 more conservative (lower LC/EC50) than for the full acute fish toxicity test. When used in
300 combination with other LoEs the TA can further support AFT conclusions. MOA assignment is
301 particularly challenging as several accepted classification systems exist. For the further
302 development of BN, the integrated consensus approach given in EnviroTox
303 (http://www.envirotoxdatabase.org/) is utilized (see Kienzler et al 2019). MOA can be useful to
304 identify compounds that are likely more toxic to fish (e.g., selected neurotoxicants) or other
305 organisms (e.g., herbicides being more toxic to algae). This can feed into the threshold approach
306 for supporting indirect LoEs and allows investigators to "make sense" of either variably toxic
307 compounds or the likelihood that a single individual assay is correct. Newer LoEs such as
308 gathering knowledge of biochemical and metabolic pathways or behavioral/physiological assays
309 can be used to inform the potential of metabolism (often reducing internal toxic burdens for a
310 compound), metabolic activation (making a compound more toxic potentially), and whether
311 these fit into existing MOA frameworks.

312 Here, we have explored more systematically the importance of the number of LoEs in data-rich
313 vs. data-poor scenarios using two example substances (Figure 6): Triclosan (high toxicity to fish)
314 and tetradecyl sulphate (medium toxicity). For each LoE, the BN model uses either evidence
315 (data) for the given substance or a prior probability distribution based on all training data,
316 depending on the scenario. In the example with triclosan, all scenarios with evidence for four or
317 three LOEs give the correct model prediction as indicated by high quality AFT results. When
318 evidence for only two LoEs are used, the toxicity level is underestimated in 2 out of 6 cases.
319 When evidence for only one LoE is used, toxicity is even more strongly underestimated in 2 out
320 of 4 cases. The BN predicts the correct interval for all scenarios where embryo data are used. In
321 the case of tetradecyl sulfate, the prediction accuracy (correct interval have highest probability)
322 increases with the number of LoEs with evidence. Compared to triclosan, the predictions for
323 tetradecyl sulfate tend to have lower precision (the peak probabilities don't get as high) and a bit
324 lower accuracy (only 8 out of 15 have correct interval, compared to 11 out of 15 for triclosan).

This is likely related to the fact that more toxicity values for algae (n=4) and Daphnia (n=6) are available for triclosan.

To generalize from these two examples, inclusion of evidence from a larger number of LoEs in this BN model seems to improve both accuracy (more cases of correctly predicted toxicity interval) and precision (higher probability of the predicted most probable interval). We also expect that inclusion of more replicate values (repetitive tests on the same endpoints) within an LoE would give more precise AFT predictions (that is, narrower probability distributions). This aspect will be further analyzed in the current SWiFT project. Actual AFT data is used within the model evaluation. Of course, this assumes that the AFT data is precise and accurate, therefore, an effort to assess reliability of toxicity tests and curate detailed results is also part of the effort. It is important to recognize that the BN model utilizes all data to understand variance in measured endpoints. This is an improvement over the common practice of using geometric mean values to summarize multiple measurements for the same taxon (Stephan et al. 1985; ECHA 2008).

Figure 5 displays how LoEs are combined in the BN model. Within each LoE (pathway), toxicity to fish is predicted from one or more nodes. The toxicity nodes are arbitrarily discretized into 5 toxicity levels, based on intervals of effect concentration (LC50): very low (>100 mg/L), low (5–100 mg/L), medium (0.5–5 mg/L), high (0.5–0.01 mg/L), and very high (<0.01 mg/L). These toxicity levels are relevant since aquatic toxicity not only used as environmental risk assessment inputs for PNEC (Predicted No Effect Concentration for ecosystems) determination, but also for classification and labelling under the Globally Harmonized System. For all parent nodes (with no incoming arrows), prior probability distributions were calculated based on all substances in the dataset. Each child node has a conditional probability table (CPT) that is used for calculating the posterior probability distribution dependent on the probability distribution of the parent node, when the model is run for a given chemical. The CPTs are constructed both from expert judgement (e.g., assignment of chemical categories) and from empirical relationships (e.g., frequency distributions of toxicity levels in different chemical categories).

Table 1 (inputs for the BN) can be thought of in a tiered approach quite easily. Direct LoE provide the best evidence for AFT, assuming a high-quality assay is performed. The more LoE

354 employed, the higher the tier. Outputs are provided along the way as a probability for a particular
355 outcome. This is a similar process to the historical tiering of aquatic toxicity data given in Cairns
356 and Dickson (1978) and described in Menzie et al. (2021). The more LoEs that are implemented
357 and the more evidence within each LoE that is gathered, both the accuracy and precision of the
358 prediction should ideally improve. As with conventional risk assessment paradigms, the level of
359 uncertainty that can be tolerated in a given environmental decision is related to the magnitude of
360 the difference between hazard and exposure concentrations and the precision required for
361 acceptance.

362 The AFT BN model is a quantitative approach to WoE (while many others are qualitative, see
363 Suter et al. 2017). The BN can conceivably be mapped to a WoE framework. The weighting in a
364 traditional WoE could be implemented as probability distributions in a BN, both within
365 individual LoEs and for integrating the LoEs (Moe et al. 2021). The assignment of conditional
366 probabilities to different variables within in a LoE in the BN corresponds to setting weights to
367 pieces of evidence in a WoE (Moe et al. 2021). Assigning a lower weight to a piece of evidence
368 can be obtained by setting wide probability distributions (representing high uncertainty or
369 variability) in the relationship from this node to its child node. Within a line of evidence the
370 calculation of posterior probabilities for the last child node of the line then accumulates the
371 weights given to all parent nodes in this line. It follows that when using the BN to predict the
372 toxicity to juvenile fish from all LoE for a chemical of concern, the weighting of the total
373 evidence for each hypothesis in a WoE (Suter et al., 2017a) can correspond to calculating the
374 posterior probability of each toxicity level ("very low", "low" etc.) in the BN.

375 **What a successful WoE for AFT Looks Like**

376 Regulatory acceptance of animal alternative approaches is challenging. There are many reasons
377 for this including the structure of regulations which vary across the globe, the dependence on
378 vertebrate data in historical assessments which could be "re-visited" if reliance on alternatives
379 take hold, the uneven levels of expertise in the scientific and regulatory communities, and fast-
380 paced evolving regulatory changes which "move, and often raise, the bar". Due to these and
381 many other factors, a successful WoE for AFT would possess the following attributes:

382 1. The approach would have global applicability;

2. Outcomes would be met with regulatory acceptance and not require significant changes to legislations or federal law;
3. Would be relevant equally for all fish species whose use varies across regulatory jurisdictions (e.g., Europe employs zebrafish most frequently, the US uses fathead minnow, Canada and the UK rely heavily upon rainbow trout, Japan uses Medaka preferentially and movement exists for China to rely upon the Chinese rare minnow).
4. The model leaves open for expansion of knowledge regarding chemical domain, new LoEs, and underlying input data;
5. Would include a renewed recognition that the "gold standard" AFT assay, is not necessarily more precise and accurate, thus, improving the transparency and curation of accepted highest quality AFT data is critical;
6. Would improve the environmental hazard and risk assessment of chemicals;
7. Be publicly available, but controlled, to ensure transparent application.

Beyond AFT, other areas of fish hazard assessment could benefit from a similar approach (and some are already in progress). Bioaccumulation assessment is identified in ECHA (2008, 2017) as being a WoE process and is discussed in more detail within this special issue by Arnot and colleagues (Arnot et al. 2021, in this Special Series) and previously by Lillicrap et al. (2016). Endocrine disruption and long-term fish toxicity were identified in OECD (2012) as also requiring WoE development. Bioaccumulation, endocrine disruption (also discussed by Mihaich and Burgoon et al. (2021) (in this Special Series) and long-term fish toxicity all involve assessment of complex endpoints that require deep expertise. A transparent, accepted WoE process for all would benefit the regulatory environment in addition to the early development of new chemical entities. Furthermore, animal alternative assays will play a pivotal role in future WoE approaches as already identified by Volz et al. 2011 and Villeneuve et al. 2014 for long-term fish toxicity, Nichols et al. (2018) for bioaccumulation and Scholz et al. (2014) for potential endocrine disruption. In all of these, the guiding principles for WoE will be the same but a much wider net of lines of evidence will likely be needed. Whether Bayesian approaches would fulill some of these needs is yet to be answered.

Many remaining challenges exist, including how to deal with novel chemicals which may not have much context for their use, MOA and properties. That said, the surface has been scratched

with respect to existing compounds whose assessments are part of the historical record, but lack animal alternative data. Compounds with difficult solubility, sorptivity, analytical verification and toxic properties will also remain challenging. Importantly, regulators and scientists will need to possess courage to make decisions when data are less than perfect (which will always be the situation) and develop new regulatory inertia in the face of a lack of precedents. Finally, since environmental risk assessment frameworks and chemical legislations are principally employed for the protection of the environment, hazard assessment strategies should be moving away from relying on acute fish toxicity for classification purposes and focus more on mechanistic/sublethal endpoints to identify chemicals of concern.

Literature Cited

Arnot JA, Toose L, Armitage JM, Embry M, Sangion A, Hughes L. 2021. A Tiered Weight-of-Evidence Approach for Bioaccumulation Assessment paper. *Integr Environ Assess Manag* (this issue).

Belanger SE, Rawlings JM, Carr GJ. 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

Brill JL, Belanger SE, Barron MG, Beasley A, Connors KA, Embry M, Carr GJ. 2021. Derivation of algal acute to chronic ratios for use in chemical toxicity extrapolation. *Chemosphere*. In Press.

Birke A, Scholz S. 2019. Zebrafish embryo and acute fish toxicity test show similar sensitivity for narcotic compounds. *ALTEX* 36:131-135.

Burden N, Benstead R, Benyon K, Clook M, Green C, Handley J, Harper N, Maynard SK, Mead C, Pearson A, Ryder K, Sheahan D, van Egmond R, Wheeler JR, Hutchinson TH. 2020. Key Opportunities to Replace, Reduce, and Refine Regulatory Fish Acute Toxicity Tests. Environ Toxicol Chem. 2020 39:2076-2089. doi: 10.1002/etc.4824

Busquet F, Strecker R, Rawlings JM, Belanger SE, Braunbeck T, Carr GJ, Cenijn P, Fochtman P, Gourmelon A, Hübler N, Kleensang A, Knöbel M, Kussatz C, Legler J, Lillicrap A, Martínez-Jerónimo F, Polleichtner C, Rzodeczko H, Salinas E, SchneidermKE, Scholz S, van den Brandhof EJ, van der Ven LTM, Walter-Rohde S, Weigt S, Witters, Halder M. 2014. OECD validation study to assess intra- and inter-laboratory reproducibility of the zebrafish embryo toxicity test for acute aquatic toxicity testing. *Reg Toxicol Pharmacol* 69:496-511.

Cairns J Jr, Dickson KL, Maki AW, Eds. (1978) Estimating the hazard of chemical substances to aquatic life. ASTM Special Technical Publication 657. Philadelphia, Pa.: American Society for Testing and Materials.

448 Carr GJ, Bailer AJ, Rawlings, JM, Belanger SE. 2018. On the impact of sample size on median
449 lethal concentration estimation in acute fish toxicity testing: Is n = 7/group enough?
450 *Environ Toxicol Chem* 37, 1565-1578. doi:10.1002/etc.4098

451 DTU Food, National Food Institute. 2018. Danish QSAR Database. User Manual for the Danish
452 (Q)SAR Database. 38p. http://qsardb.food.dtu.dk/DB_user_manual_21_12_2018.pdf.

453 European Commission Joint Research Center (EC JRC). 2014. JRC QSAR Model Database User
454 Manual, Version 2. Ispra, Italy. 33p

455 ECHA (European Chemicals Agency). 2008. Guidance on information requirements and
456 chemical safety assessment Chapter R.10: Characterisation of dose [concentration]-
457 response for environment. May 2008, Helsinki, Finland. 65 p.
458 https://echa.europa.eu/documents/10162/13632/information_requirements_r10_en.pdf/bb9
459 02be7-a503-4ab7-9036-d866b8ddce69

460 ECHA. 2017. Expert Workshop on the potential regulatory application of the Fish Embryo Acute
461 Toxicity (FET) Test under REACH, CLP and the BPR, 3-4 May 2017 in Helsinki.

462 European Commission. 2020. Report on the statistics on the number of animals used for
463 experimental and other scientific purposes in the member states of the European Union in
464 2015-2017. Commission staff working document. https://eur-lex.europa.eu/legal-
465 content/EN/TXT/?qid=1581689520921&uri=CELEX:52020SC0010

466 EFSA Scientific Committee, Hardy, A., Benford, D., Halldorsson, T., Jeger, M.J., Knutsen,
467 H.K., More, S., Naegeli, H., Noteborn, H., Ockleford, C., Ricci, A., Rychen, G., Schlatter,
468 J.R., Silano, V., Solecki, R., Turck, D., Benfenati, E., Chaudhry, Q.M., Craig, P.,
469 Frampton, G., Greiner, M., Hart, A., Hogstrand, C., Lambre, C., Luttik, R., Makowski, D.,
470 Siani, A., Wahlstroem, H., Aguilera, J., Dorne, J.-L., Fernandez Dumont, A., Hempen, M.,
471 Valtue~na Martínez, S., Martino, L., Smeraldi, C., Terron, A., Georgiadis, N., Younes, M.,
472 2017. Guidance on the use of the weight of evidence approach in scientific assessments.
473 EFSA J. 15 (8), e04971.

474 Fischer M, Belanger SE, Berckmans P, Bernhard MJ, Bláha L, Coman Schmid DE, Dyer SD,
475 Haupt T, Hermens JLM, Hultman MT, Laue H, Lillicrap A, Mlnaříková M, Natsch A,
476 Novák J, Sinnige TL, Tollefsen KE, von Niederhäusern V, Witters H, Županič A, Schirmer
477 K. 2019. Repeatability and reproducibility of the RTgill-W1 cell line assay for predicting
478 fish acute toxicity. *Toxicolog Sci* 169:353–364.

479 Hall T, Martin O, Belanger S, Galay-Burgos M, Guiney PD, Maack G, Stubblefield W. 2017.
480 Weight of evidence for ecotoxicological effects characterization in regulatory decision-
481 making. *Integr Environ Assess Manag* 13:573-579.

482 Hutchinson TH, Barrett S, Buzby M, Constable D, Hartmann A, Hayes E, Huggett D, Laenge R,
483 Lillicrap AD, Straub JO, Thompson RS. 2003. A strategy to reduce the numbers of fish
484 used in acute ecotoxicity testing of pharmaceuticals. *Environ Toxicol Ch*em 22:3031-3036.
485 https://doi.org/10.1897/02-558.

ISO. 2019. Water quality — Determination of acute toxicity of water samples and chemicals to a fish gill cell line (RTgill- W1). ISO 21115:2019. doi:10.3403/30361171u

Katsiadaki I, Ellis T, Andersen L, Antczak P, Blaker E, Burden N, Fisher T, Green C, Labram B, Pearson A, Petersen K, Pickford D, Ramsden C, Rønneseth A, Ryder K, Sacker D, Stevens C, Watanabe H, Yamamoto H, Sewell F, Hawkins P, Rufli H, Handy RD, Maynard SK, Jacob MN. 2021 Dying for change: A roadmap to refine the fish acute toxicity test after 40 years of applying a lethal endpoint. *Ecotoxicol Environ Safety* 223 (2021) https://doi.org/10.1016/j.ecoenv.2021.112585

Kienzler A, Barron MG, Belanger SE, Beasley A, Embry MR. 2017. Mode of Action (MOA) assignment classifications for ecotoxicology: An evaluation of Approaches. *Environ Sci Technol* 51:10203-10211.

Klüver N, Ortmann J, Paschke H, Renner P, Ritter AP, Scholz A.. 2014. Transient overexpression of adh8a increases allyl alcohol toxicity in Zebrafish embryos. *PLoS One* 9:e90619. doi: 10.1371/journal.pone.0090619.

Klüver N, Konig M, Ortmann, J, Massei R, Paschke A, Kühne R, Scholz S. 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

Lillicrap A, Belanger S, Burden N, Du Pasquier D, Embry MR, Halder M, Lampi MA, Lee L, Norberg-King T, Rattner BA, Schirmer K, Thomas P. 2016. Alternative approaches for vertebrate ecotoxicity tests in the 21st century: a review of the current status and developments over the last two decades. *Environ Toxicol Chem* 35:2637-2646.

Lillicrap A, Springer T, Tyler CR. 2016b. A tiered assessment strategy for more effective evaluation of bioaccumulation of chemicals in fish. Regulatory Toxicology and Pharmacology, Volume 75, 20-26 https://doi.org/10.1016/j.yrtph.2015.12.012.

Lillicrap A, Moe J, Connors K, Landis W, Madsen A, Rawlings J, Belanger S. 2020. Development of a Bayesian network to strengthen weight of evidence for fish embryo toxicity data to predict acute fish toxicity. *Integr Environ Assess Manag* 16(4):452–460.

Lammer E, Carr GJ, Wendler K, Rawlings JM, Belanger SE, Braunbeck T. 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

Landis WG. 2021. The origin, development, application, lessons learned, and future regarding the Bayesian network relative risk model for ecological risk assessment. *Integr Environ Assess Manag* 17(1): 79– 94. https://doi.org/10.1002/ieam.4351

Loerracher AK, Braunbeck T. 2021. Cytochrome P450-dependent biotransformation capacities in embryonic, juvenile and adult stages of zebrafish (Danio rerio)—a state-of-the-art review. Arch Toxicol 95:2299–2334.

Martin O, Adams J, Beasley A, Belanger S, Breton RL, Brock TCM, Buonsante V, Galay Burgos M, Green J, Guiney PD, Hall T, Hanson M, Harris M, Henry TR, Huggett D, Junghans M, Kivi M, Laskowski R, Maack G, Moermond CTA, Panter G, Pease A, Poulsen V, Roberts M, Rudén C, Schlekat CE, Schoeters I, Solomon KR, Staveley J, Stubblefield B, Sumpter JP, St. J. Warne M, Wentsel R, Wheeler JR, Wolff B, Yamazaki K, Zahner H, Ågerstrand M. 2019. Improving environmental risk assessments of chemicals: steps towards evidence-based ecotoxicology. *Environ Internat* 128(2019):210-217.

Menzie CA, Belanger SE, Guiney PD. 2021. A Recommended Framework for the Practical Application and Advancement of the Use of Weight-of-Evidence for Environmental Risk Assessment. *Integr Environ Assess Manag* (this issue).

Moe SJ, Madsen AL, Connors KA, Rawlings JM, Belanger S, Landis WG, Wolf R, Lillicrap AD. 2020. Development of a hybrid Bayesian network model for predicting acute fish toxicity using multiple lines of evidence. *Environ Model Software* 126:104655. https://doi.org/10.1016/j.envsoft.2020.104655.

Moe, S.J., Carriger, J.F. and Glendell, M. 2021a. Increased Use of Bayesian Network Models Has Improved Environmental Risk Assessments. *Integr Environ Assess Manag* 17: 53-61. https://doi.org/10.1002/ieam.4369

Moe SJ, Wolf R, Xie L, Landis WG, Kotamäki N, Tollefsen KE. 2021b. Quantification of an adverse outcome pathway network by Bayesian regression and Bayesian network modeling. *Integr Environ Assess Manag* 17(1):147–164. https://doi.org/10.1002/ieam.4348

Natsch A, Laue H, Haupt T, von Niederhausern V, Sanders G. 2018. Accurate prediction of acute fish toxicity of fragrance chemicals with the RTgill-W1 cell assay. *Environ Toxicol Chem* 37:931-941.

Nichols J, Fay K, Bernhard MJ, Bischof I, Davis J, Halder M, Hu J, Johanning K, Laue H, Nabb D, Schlechtriem C, Segner H, Swintek J, Weeks J, Embry M. 2018. Reliability of in vitro methods used to measure intrinsic clearance of hydrophobic organic chemicals by rainbow trout: results of an international ring trial. *Toxicol Sci* 164:563–575. doi: 10.1093/toxsci/kfy113

Norberg-King TJ, Hughes SA, Embry MR, Belanger SE, Braunbeck T, Butler JD, Dorn PB, Farr B, Guiney PD, Jeffries M, Journel R, Leonard ML, McMaster M, Oris JT, Ryder K, Segner H, Senac T, Van Der Kraak G, Wilson P, Whale G. 2018. An international perspective on the tools and concepts for effluent toxicity assessments in the context of animal alternatives. *Environ Toxicol Chem* 37(11):2745-2757.

[OECD] Organisation for Economic Co-operation and Development. 2004. Daphnia sp. acute immobilisation test. Guidelines for the Testing of Chemicals 202. doi:10.1787/9789264069947-en.

[OECD] Organisation for Economic Co-operation and Development. 2004. The Report from the Expert Group on (Quantitative) Structure-Activity Relationships [(Q)SARs] on the Principles for the Validation of (Q)SARs. Monograph Number 49. 206p.

[OECD] Organisation for Economic Co-operation and Development.2005. Guidance document on the validation and international acceptance of new or updated test methods for hazard assessment. OECD Series on Testing and Assessment, Number 34. Paris, France. 96p.

[OECD] Organisation for Economic Co-operation and Development. 2010. Short guidance on the threshold approach for acute fish toxicity. Series of Testing & Assessment 126. doi:10.1787/9789264078413-en.

[OECD] Organisation for Economic Co-operation and Development. 2011b. Freshwater Alga and Cyanobacteria, Growth Inhibition Test. Guidelines for the Testing of Chemicals 201. doi:10.1787/20745761 doi:10.1787/9789264069923-en

[OECD] Organisation for Economic Co-operation and Development. 2012a. Fish testing framework. OECD Series on Testing and Assessment, Number 171. doi:10.1787/9789264221437-en.

[OECD] Organisation for Economic Co-operation and Development. 2013. Fish Embryo Acute Toxicity (FET) Test. Test Guidelines for the Testing of Chemicals 236. doi:10.1787/9789264203709-en.

[OECD] Organisation for Economic Co-operation and Development. 2018a. Determination of in vitro intrinsic clearance using cryopreserved rainbow trout hepatocytes (RT-HEP). 319A Test Guidelines for the Testing of Chemicals. https://doi.org/10.1787/9789264303218-en

[OECD] Organisation for Economic Co-operation and Development. 2018b. Determination of in vitro intrinsic clearance using rainbow trout liver S9 sub- cellular fraction (RT-S9). 319B Test Guidelines for the Testing of Chemicals. https://doi.org/10.1787/2074577x

[OECD] Organisation for Economic Co-operation and Development. 2019b. Fish, Acute Toxicity Testing. 203 Test Guidelines for the Testing of Chemicals. doi:10.1787/20745761-en.

[OECD] Organisation for Economic Co-operation and Development. 2020. Toolbox v.4.4 - User manual, QSAR Toolbox, The OECD QSAR Toolbox for Grouping Chemicals into Categories. Document Version 1.4, February 2020. 26p. https://qsartoolbox.org/wp-content/uploads/2020/02/Toolbox-4.4-Installation-Manual.pdf

OECD. 2021. Fish Cell Line Acute Toxicity: The RTgill-W1 cell line assay. Test Guidelines for the Testing of Chemicals 249. https://doi.org/10.1787/c66d5190-en.

595 Paparella M, Scholz S, Belanger S, Bichere P, Braunbeck T, Connors K, Faßbender C, Halder
596 M, Lillicrap A, Liska R, Schirmer K, Stoddart G, Thomas P, Walter-Rhode S. 2021.
597 Limitations and uncertainties of acute fish toxicity assessments are reducible by
598 alternatives. *ALTEX* 38:20-32. *doi:10.14573/altex.2006051*

599 Rawlings JM, Belanger SE, Connors KA, Carr GJ. 2019. Fish embryo tests and acute fish
600 toxicity tests are interchangeable in the application of the threshold approach. *Environ*
601 *Toxicol Chem* 38:671-681. doi:10.1002/etc.4351

602 Russom CL, Bradbury S, Broderius SJ, Hammermeister DE, Drummond RA. 1997. Predicting
603 modes of toxic action from chemical structure: Acute toxicity in the fathead minnow
604 (*Pimephales promelas*). *Environ Toxicol Chem* 16:948–967.

605 Schirmer K, Tanneberger K, Kramer NI, Völker D, Scholz S, Hafner C, Lee LEJ, Bols NC,
606 Hermens JLM. 2008. Developing a list of reference chemicals for testing alternatives to
607 whole fish toxicity tests. *Aquatic Toxicol* 90:128-137.

608 Scholz S, Renner P, Ortego LS, Belanger S, Busquet F, Davi R, Demeneix B, Denny J, Léonard
609 M, McMaster M, Villeneuve D, Embry M. 2013. Alternatives to in vivo tests to detect
610 endocrine disrupting chemicals (EDCs) in fish and amphibians. *Critical Rev Toxicol*
611 43:45-72.

612 Sobanska M, Scholz S, Nyman A-M, Cesnaitis R, Gutierrez Alonso S, Kluver N, Kuhne R, Tyle
613 H, de Knecht J, Dang Z, Lundbergh I, Carlon C, De Coen W. 2018. Applicability of the
614 fish embryo acute toxicity (FET) test (OECD 236) in the regulatory context of Registration,
615 Evaluation, Authorisation, and Restriction of Chemicals (REACH). *Environ Toxicol*
616 *Chem*. 37:657–670.

617 Strähle U, Scholz S, Geisler R, Greiner P, Hollert H, Rastegar S, Schumacher A, Selderslaghs I,
618 Weiss C, Witters H, Braunbeck T. 2012. Zebrafish embryos as an alternative to animal
619 experiments—A commentary on the definition of the onset of protected life stages in
620 animal welfare regulations. *Repro Toxicol* 33:128–132.

621 Suter G, Cormier S, Barron M. 2017a. A weight of evidence framework for environmental
622 assessments: inferring qualities. *Integr Environ Assess Manag* 13:1038–1044.

623 Suter, G., Cormier, S., Barron, M., 2017b. A weight of evidence framework for environmental
624 assessments: inferring quantities. *Integr Environ. Assess. Manag* 13:1045–1051.

625 Tannenberger K, Knöbel M, Busser FJM, Sinnige TL, Hermens JLM, Schirmer K. 2013.
626 Predicting fish acute toxicity using a fish gill cell line-based toxicity assay. *Environ Sci*
627 *Technol* 47:1110−1119.

628 [UBA] Umweltbundesamt, German Federal Environment Agency. 2020. Development of an
629 OECD Guidance Document for the Application of OECD Test Guideline 236 (Acute Fish
630 Embryo Toxicity Test) Final Report, T. Braunbeck, S. Böhler, S. Knörr, A.-K. Lörracher,
631 K. Pelka and B. Kais. Report No. (FB) FB000024/ENG. 106p.

632 [US EPA]. United States Environmental Protection Agency. 2020. Ecological Structure Activity
633 Relationships (ECOSAR) Predictive Model. https://www.epa.gov/tsca-screening-
634 tools/ecological-structure-activity-relationships-ecosar-predictive-model. ECOSAR V. 2.0.

635 Villeneuve D, Volz DC, Embry MR, Ankley GT, Belanger SE, Léonard M, Schirmer K,
636 Tanguay R, Truong L, Wehmas L. 2014. Investigating alternatives to the fish early life-stage
637 rest: a strategy for discovering and annotating adverse outcome pathways for early fish
638 development. *Environ Toxicol Chem* 33:158-169.

639 Volz DC, Belanger S, Embry M, Padilla S, Sanderson H, Schirmer K, Scholz S, Villeneuve D.
640 2011. Adverse outcome pathways during early fish development: a conceptual framework
641 for identification of chemical screening and prioritization strategies. *Toxicol Sci* 123:349–
642 358.

643

644

645 Table 1. Strength of attributes presently incorporated or being considered in the BN for the AFT
646

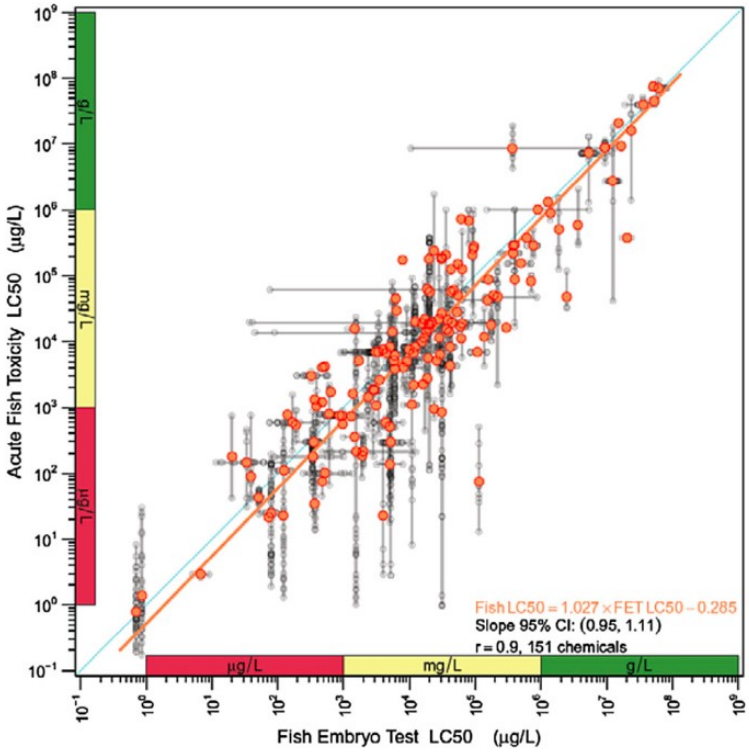
Attribute	Relevant Test Guideline(s)	Measured endpoint	Strength (potential weight or inference towards AFT)	Status in Moe et al. 2020 BN	Utility for Tiered Weight of Evidence
Physical-chemical properties	Numerous, compound dependent	Solubility, log Kow, pKa, molecular weight	Strong, influence on MOA	Included	Essential
Mode of Action	Scientific judgement	Assignment of mode	Strong	Included, see Kienzler et al. (2017) and www.EnviroToxdatabase.org	Important
Measured Acute Fish Toxicity	OECD 203	96 h LC50	Direct evidence	Used as context for similarly acting (MoA) chemicals (in the line "Chemical category"	Direct evidence, may still need to weigh conflicting data (variability in test results; use of various species)
Fish QSAR	Various accepted platforms (ECOSAR, Danish QSAR, OECD Toolbox)	96 h LC50	Strong when QSARs are validated and chemical is within domain of applicability	Included in the line "Chemical properties"	Indirect evidence
Measured Acute Fish Embryo Toxicity	OECD 236	96 h LC50	Strong	Included in the line "Embryo"	Direct evidence
Measured Acute *Daphnia* Toxicity	OECD 202	48 h EC50	Must be used in combination with other attributes; potential applicability in Threshold Approach	Included in the line "Other taxa"	Indirect evidence when used in combination with algal inhibition
Daphnia QSAR	ECOSAR, OECD	48 EC50	Strong when QSARs are	Candidate	Indirect evidence

	QSAR Toolbox		validated and chemical is within domain of applicability; must be linked by other means to fish toxicity (TA)		when used in combination with algal inhibition
Measured Algal acute inhibition	OECD 201	72 h ErC50 96 h EC50	Moderate; numerous species and complexities of regulatory endpoints *sensu* Brill et al. 2021; useful/necessary within TA	Included in the line "Other taxa"	Indirect evidence when used in combination with algal inhibition
Algae QSAR	ECOSAR, OECD QSAR Toolbox	72 h EC50 96 h EC50	Moderate; few broadly available QSARs	Candidate	Indirect evidence when used in combination with algal inhibition
Fish gill cytotoxicity	OECD 249 ISO 21115	24 h EC50	Moderate; growing database, prospects are very good	To be included	Inferential, depends on MoA
Mode of Action assignment	EnviroTox	N/A	Complexity across MoA schemes (Kienzler et al. 2017); Important to identify specifically-acting chemicals	Included in the line "Other taxa", to be expanded	Supportive
Neurotoxicant behavior assay	In development	Neural impact, Yes/No	In progress (Touch evocation assay); important means to identify this class	In exploration	Supportive

Fish metabolism	Few standardized assays; OECD 319a,b; Km QSARs	Various biochemical and physiological measurements	In development; much is known but not fully catalogued (Braunbeck et al. 2019, Loerracher and Braunbeck 2021)	In exploration	Supportive

647
648

649



650
651

652 Figure 1. Comparison of fish embryo test (FET) versus acute fish toxicity (AFT) for 151
653 compounds from Belanger et al. (2013).

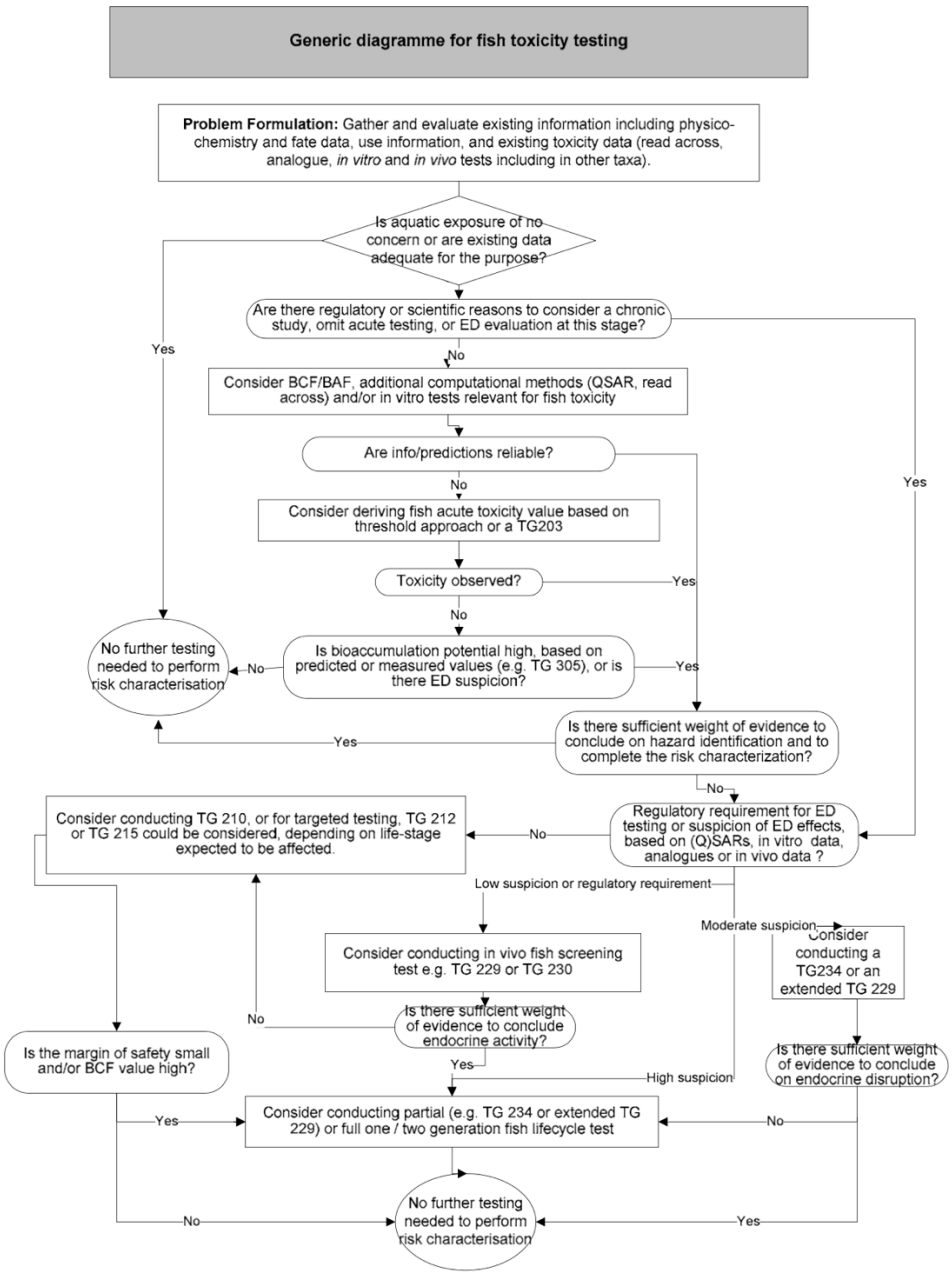
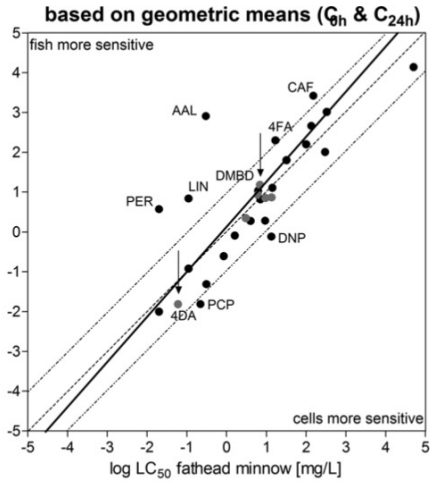


Figure 7.1: A generic testing strategy illustrating general principles of how to optimise fish toxicity testing needs. This strategy makes use of as much prior information as possible in order to determine the need for fish testing. Regulatory requirements in particular jurisdictions may require a more complex or testing-rich assessment. Risk characterization and other chemical assessments may include, for example, a PBT assessment, hazard classification, and / or endocrine assessment.

Figure 2. An integrated approach to informing the environmental hazards of chemicals to fish (Reproduced from the OECD Fish Framework (OECD 2012)).

A



B

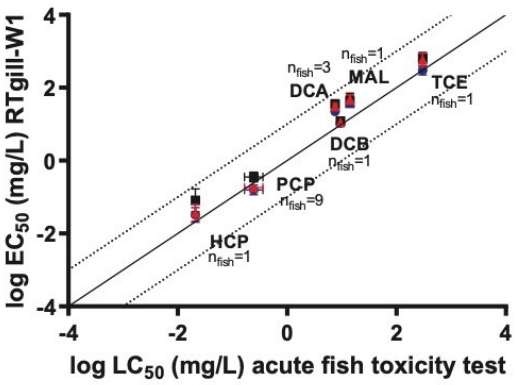
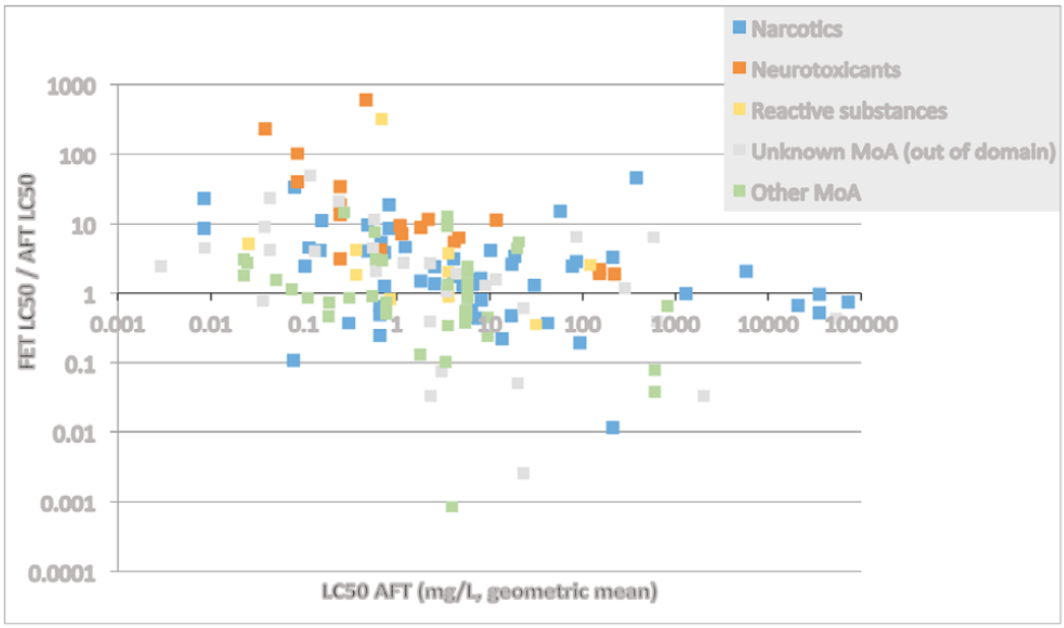


Figure 3. (A) Linear correlations between acute fathead minnow toxicity and fish cell line EC50 values for metabolic activity as given in Tannenberger et al. (2013). (B) Correspondence of round-robin study-derived average log EC50 values obtained for the 3 different cell viability measurements per test chemical with average logLC50 values from fish acute toxicity testing as given in Fischer et al. (2019). Correspondence of the cell line derived data with *in vivo* data obtained from the US EPA fathead minnow database (see Russom et al. 1997 and Tanneberger et al. 2013).



668
669 Figure 4. Differences between fish embryo toxicity (FET) and acute fish toxicity (AFT) effect
670 concentrations as described by Sobanska et al. (2018). The ratio between FET and AFT (y-axis)
671 over the range of median lethal concentrations (LC50s) in the acute fish toxicity (x-axis).
672 Different modes of action are indicated in different legend colors.
673

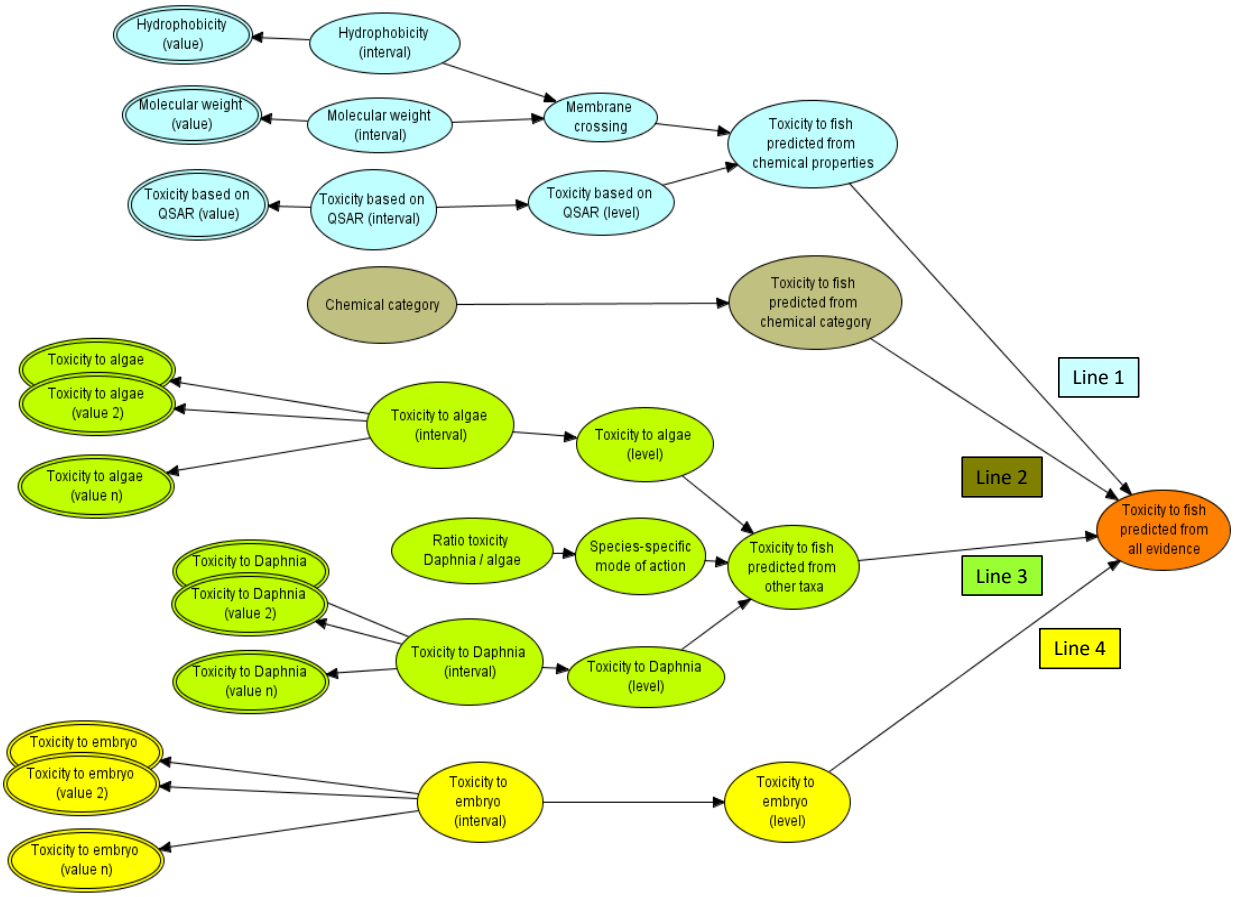


Figure 5. Conceptual version of the Bayesian network model to predict Acute Fish Toxicity from four lines of evidence (after Moe et al. 2020). Nodes with double outlines represent input values (observed toxicity values) in continuous scale; the model currently accepts up to 10 replicate observations. All other nodes have discrete states, such as concentration intervals.

A

	Triclosan				
	>100	5-100	0.5-5	0.01-0.5	<0.01
LOEs	**Very low**	**Low**	**Medium**	**High**	**Very high**
PCOE	0.008	0.078	0.240	**0.559**	0.075
PCO	0.036	0.179	0.344	**0.393**	0.048
PCE	0.012	0.017	0.289	**0.550**	0.042
POE	0.021	0.051	0.161	**0.648**	0.120
COE	0.035	0.191	0.320	**0.403**	0.051
PC	0.047	0.220	**0.361**	**0.344**	0.028
PO	0.080	0.178	0.287	**0.358**	0.096
CO	0.079	0.309	**0.349**	0.231	0.032
PE	0.032	0.095	0.209	**0.596**	0.068
CE	0.042	0.230	0.347	**0.356**	0.026
OE	0.078	0.193	0.256	**0.391**	0.082
P	0.094	0.215	0.304	**0.331**	0.056
C	0.095	**0.355**	0.344	0.189	0.017
O	0.151	**0.289**	0.282	0.207	0.072
E	0.092	0.226	0.270	**0.370**	0.042

Probability range

0.008-0.050
0.051-0.100
0.101-0.150
0.151-0.200
0.201-0.250
0.251-0.300
0.301-0.350
0.351-0.400
0.401-0.450
0.451-0.500
0.501-0.550
0.551-0.600
0.601-0.650

B

	Triclosan				
	>100	5-100	0.5-5	0.01-0.5	<0.01
LOEs	**Very low**	**Low**	**Medium**	**High**	**Very high**
PCOE	0.039	0.241	**0.420**	0.284	0.017
PCO	0.091	0.381	**0.395**	0.122	0.011
PCE	0.044	0.242	**0.405**	0.285	0.025
POE	0.079	0.229	0.282	**0.379**	0.030
COE	0.040	0.244	**0.419**	0.282	0.015
PC	0.100	0.374	**0.382**	0.128	0.016
PO	0.164	**0.348**	0.300	0.160	0.027
CO	0.092	0.384	**0.393**	0.121	0.010
PE	0.091	0.223	0.271	**0.371**	0.044
CE	0.045	0.245	**0.403**	0.283	0.023
OE	0.081	0.232	0.281	**0.379**	0.028
P	0.177	**0.336**	0.284	0.164	0.039
C	0.101	0.377	**0.380**	0.127	0.015
O	0.166	**0.351**	0.298	0.160	0.025
E	0.092	0.226	0.270	**0.370**	0.042

Figure 6. Comparison of BN model predictions of fish toxicity under data-rich and data-poor scenarios for two example substances: (a) Triclosan and (b) Tetradecyl sulphate. The 15 scenarios use available data for one, two three or all four of the lines of evidence (LOEs): P = physical/chemical properties, C = chemical category, O = other taxa (algae and *Daphnia*), E = fish embryo. For each scenario, the table shows the predicted probability distribution across the five toxicity levels, defined by intervals of LC50 values (in mg/L). The outlined column indicates the observed toxicity interval, while bold numbers highlight the interval with the highest predicted probability. Colors indicate broad ranges of probability of occurrence ranging from dark green (low) to red (high).