Dying for change: A roadmap to refine the fish acute toxicity test after 40 years of applying a lethal endpoint

Ioanna Katsiadaki [a],[*], Tim Ellis [a], Linda Andersen [b], Philipp Antczak [c],[w], Ellen Blaker [a], Natalie Burden [d], Tom Fisher [e], Christopher Green [f], Briony Labram [j], Audrey Pearson [l], Karina Petersen [m], Dan Pickford [n], Chris Ramsden [o], Anita Rønneseth [p], Kathy Ryder [q], Dominic Sacker [s], Chloe Stevens [t], Haruna Watanabe [u], Hiroshi Yamamoto [v], Fiona Sewell [j], Penny Hawkins [h], Hans Rufli [r], Richard D. Handy [g], Samuel K. Maynard [k], Miriam N. Jacobs [i]

[a] Centre for Environment, Fisheries and Aquaculture Science (Cefas), Weymouth Laboratory, Barrack road, Dorset DT4 8UB, UK
[b] ILAB, Thormøhlens gate 55, 5006 Bergen, Norway
[c] Institute of Integrative Biology, Biology, Biosciences Building, University of Liverpool, Crown Street, Liverpool L69 7ZB, UK
[d] NC3Rs, Gibbs Building, 215 Euston Road, London NW1 2BE, UK
[e] Chemical Regulation Division, Mallard House, 3 Peasholme Green, York YO1 7PX, UK
[f] Department for Environment, Food and Rural Affairs, London, UK
[g] School of Biological and Marine Sciences, University of Plymouth, Plymouth, Devon PL4 8AA, UK
[h] Animals in Science Department, Royal Society for the Prevention of Cruelty to Animals, Wilberforce Way, Southwater, Horsham, West Sussex RH13 9RS, UK
[i] Department of Toxicology, Centre for Radiation, Chemical and Environmental, Hazards Public Health England, Chilton, Oxfordshire OX11 0RQ, UK
[j] NC3Rs, UK
[k] AstraZeneca, Global Sustainability, Eastbrook House, Shaftesbury Road, Cambridge CB2 8DU, UK
[l] Environment Agency, Chemical Assessment Unit, Red Kite House, Howbery Park, Wallingford, Oxfordshire OX10 8BD, UK
[m] NIVA, Gaustadalléen 21, NO-0349 Oslo, Norway
[n] Syngenta Ltd, Jealott's Hill International Research Station, Bracknell RG42 6EY, UK
[o] AgroChemex Environmental, Aldhams Research Farm, Dead Lane, Lawford, Manningtree, Essex CO11 2NF, UK
[p] Department of Biological Sciences, University of Bergen, P.O. Box 7803, N-5006 Bergen, Norway
[q] Northern Ireland Animals (Scientific Procedures) Act (ASPA) Inspectorate, Room C4.3, Castle Buildings, Stormont Estate, Belfast BT4 3SQ, UK
[r] ecotoxsolutions, Unterer Rheinweg 114, CH-4057 Basel, Switzerland
[s] Covance CRS Research Limited, Shardlow Business Park, London Road, Shardlow DE72 2GD, UK
[t] RSPCA, UK
[u] National Institute for Environmental Studies (NIES), 16-2 Onogawa, Tsukuba-City, Ibaraki 305-8506 Japan
[v] NIES, Japan
[w] Center for Molecular Medicine Cologne, University Hospital Cologne, 50931 Cologne, Germany

ARTICLE INFO

ABSTRACT

The fish acute toxicity test (TG203; OECD, 2019) is frequently used and highly embedded in hazard and risk assessment globally. The test estimates the concentration of a chemical that kills 50% of the fish (LC_{50}) over a 96 h exposure and is considered one of the most severe scientific procedures undertaken. Over the years, discussions at the Organisation for Economic Co-operation and Development (OECD) have resulted in changes to the test which reduce the number of fish used, as well as the development of a (potential) replacement test (TG236, OECD, 2013). However, refinement of the mortality endpoint with an earlier (moribundity) endpoint was not considered feasible during the Test Guideline's (TG) last update in 2019. Several stakeholders met at a UK-based workshop to discuss how TG203 can be refined, and identified two key opportunities to reduce fish suffering: (1) application of clinical signs that predict mortality and (2) shortening the test duration. However, several aspects need to be addressed before these refinements can be adopted. TG203 has required recording of major categories of sublethal clinical signs since its conception, with the option to record more detailed signs introduced in the 2019 update. However, in the absence of guidance, differences in identification, recording and reporting of clinical signs between technicians and laboratories is likely to have generated piecemeal data of varying quality.

* Corresponding author.
 E-mail address: Ioanna.katsiadaki@cefas.co.uk (I. Katsiadaki).

Harmonisation of reporting templates, and training in clinical sign recognition and recording are needed to standardise clinical sign data. This is critical to enable robust data-driven detection of clinical signs that predict mortality. Discussions suggested that the 96 h duration of TG203 cannot stand up to scientific scrutiny. Feedback and data from UK contract research organisations (CROs) conducting the test were that a substantial proportion of mortalities occur in the first 24 h. Refinement of TG203 by shortening the test duration would reduce suffering (and test failure rate) but requires a mechanism to correct new results to previous 96 h LC_{50} data. The actions needed to implement both refinement opportunities are summarised here within a roadmap. A shift in regulatory assessment, where the 96 h LC_{50} is a familiar base for decisions, will also be critical.

1. Introduction

The fish acute toxicity test is a frequently used test for chemical hazard characterisation (Burden et al., 2017, 2020) and is strongly embedded in the risk assessment of chemicals globally. In brief, fish are exposed to the test chemical for a period of 96 h, mortalities and abnormalities are recorded at 24, 48, 72 and 96 h and the concentration which kills 50% of the fish (LC_{50}) is derived. The test was originally adopted as an Organisation for Economic Cooperation and Development (OECD) test guideline (TG203) in 1981. At that time, there was little consideration of animal welfare such as managing the health of stock animals or addressing the 3Rs (Replacement, Reduction, Refinement; Russel and Burch, 1959) within the test protocol. Over the past 40 years, TG203 has been revised on three occasions (1984, 1992 and 2019). These updates reflected scientific and regulatory needs, harmonisation with acute toxicity tests used by OECD member countries (e.g. US EPA Test OCSPP 850.1075) and recommendations from the Fish Toxicity Testing Framework (FTTF, OECD, 2014). The FTTF encouraged consideration of reliable alternative methods (QSAR, read-across, fish embryos, fish cell lines and others) using a weight-of-evidence approach before conducting the test. However, discussions at the OECD validation management group for ecotoxicity (VMG-Eco), which oversees OECD ecotoxicity test guideline development and revision, concluded that fish acute toxicity data continue to be core regulatory requirements and current suggested alternative methods (e.g. TG236, OECD, 2013) are not sufficient for all jurisdictions and testing needs, including REACH (ECHA (European Chemicals Agency), 2016) Another key FTTF recommendation (OECD, 2014) was to address animal welfare, a point of concern in TG203. Therefore, given that replacement is not an (imminent) option, efforts during the 2019 TG203 update, focused on reducing the level of suffering that fish experience during the test.

Pollution is considered as one of the top direct drivers of biodiversity loss (IPBES, 2019) and effective chemicals management internationally is an important regulatory measure to help reduce environmental harm. Chemical pollution is increasingly viewed as a planetary boundary issue and lessons learned from the past, where highly toxic chemicals were released in the environment, indicate that hazard and risk assessment for chemicals must remain rigorous and regulation aligned with the best available science. Fish represent a significant phylogenetic group of animals that we aim to protect and the LC_{50} - as derived from TG203 for four decades - has become embedded within chemical hazard and risk assessments. The increasing size of the human population has been a key driver in the growth of the chemicals industry, where sales are expected to double between 2017 and 2030 (UNEP, 2019) to meet societal demands for key sustainable development goals such as zero hunger and good health. Certain chemicals, such as plant protection products, are key in achieving these goals but do present acute toxicological concern. This landscape creates the need to further consider how we can meet the societal standards on animal welfare whilst maintaining the environmental protection goal that fish acute toxicity assessment provides.

One way to reduce suffering during animal experiments is by applying earlier, more humane endpoints, including, for example, moribundity. Early termination of fish showing signs of "considerable suffering" is featured in other test guidelines such as the Medaka Extended One Generation Test (MEOGRT; TG240; OECD, 2015).

However, generally these are guidelines designed to examine less severe sub-lethal toxic effects, not acute toxicity. Historically, acute toxicity assessment has employed a lethality estimate, the LC_{50}. However, although relevant signs of toxicity for fish have been suggested (Rufli, 2012), no consensus was reached in the OECD VMG-Eco on the effectiveness of these signs in predicting death in acute fish testing during the most recent TG203 update. An early stumbling block was the lack of agreement on a definition of moribund fish. However, it has since emerged that a key obstacle in moving away from mortality in fish acute toxicity studies is concern that such change would result in a reduction in the derived LC_{50}, which in turn would lead to an overly conservative risk assessment and increase the use of animals by triggering further tests. For example, when assessing a pesticide active substance for use in plant protection products, if a risk is identified and driven by fish, options for toxicity refinement include use of the geometric mean and species sensitivity distributions. Both these options would involve the acute testing of up to 4 additional species of fish (EFSA (European Food Safety Authority), 2013), noting vertebrate studies must be justified under EU regulations. Other assessment paradigms may result in the triggering of chronic testing where reduced toxicity values result from implementation of early clinical signs. It is therefore critical that the early signs need to be genuinely predictive of death, a link that is still elusive. Consequently, mortality remains the endpoint in the fish acute toxicity test.

In contrast, the mammalian acute oral toxicity test (OECD TG401) was deleted in 2002 due to animal welfare concerns (OECD, 2002), and superseded by refined methods that also reduced animal use. Refined methods have also been developed for other acute mammalian and avian toxicity tests, including inhalation (Sewell et al., 2015, 2018) and dermal routes of administration. Improvements involved the adoption of the 'up-and-down' procedure and the use of 'evident toxicity', whilst for avian testing sequential testing methods have been employed. The up-and-down procedure for mammalian testing is based on a range estimate, rather than a point estimate of the LC50 that requires the use of multiple doses/concentrations to construct a concentration/dose-mortality curve. This allowed a significant reduction in the number of animals used for classification and labelling purposes (e.g. OECD TG423, 425, 436 for acute oral and acute inhalation toxicity studies). For avian testing it has been demonstrated that the use of a similar sequential testing methodology (OECD 223) resulted in a reduction of around 65% in animal usage (Maynard et al., 2014; Edwards et al., 2017) for a pesticide data set. Another approach, i.e. the use of the threshold approach (OECD, 2010a) or the limit test, has been adopted in TG203 (OECD, 2019a, 2019b) and is expected to reduce the number of fish used, (Jeram et al., 2005; Creton et al., 2014). In the limit test, only one group of fish is exposed for 96 h to 100 mg L^{-1} or at the limit of solubility of the test chemical. In the absence of mortality, there is at least 99% confidence that the LC_{50} is greater than the tested concentration and would be considered non-toxic to fish for most safety assessment purposes (OECD, 2019a, 2019b). In the "threshold approach" - an initial test is carried out using just one concentration selected based on the results of prior *Daphnia* and algae toxicity tests. Only if mortality is observed at the threshold concentration a full TG203 (at least five concentrations) is triggered.

A key refinement method in mammalian testing is the use of evident

toxicity in place of death as the endpoint[1]: the range estimate of the LC_{50}/LD_{50} is predicted based on clear signs of toxicity at a lower dose/concentration, using biometrical evaluation tools and/or observations of clinical signs to assess evident toxicity in individual animals as an endpoint (OECD TG402, 433, Mielke et al., 2017; Sewell et al., 2015, 2018). This refinement reduces animal suffering and was accepted in the revised acute oral toxicity (TG420) in 2002 and in the revised inhalation (TG433) and dermal routes (TG402) in 2017. The TG203 (2019) update included a comprehensive list of fish clinical signs (along with definitions and a suggested score sheet), to aid systematic recording of the 'observations' required in TG203, ultimately allowing their retrospective evaluation as predictors of death of that same animal. This approach has some similarities with the concept of evident toxicity adopted in mammalian acute tests, although there is a subtle, yet important difference. In fish there is an explicit requirement to assess certainty of death at individual animal level, whilst in mammalian testing evident toxicity is the identification of signs in a group of animals that will predict animals will die if exposed to *higher* levels, it is not a signal for euthanasia and not equivalent to death of the animal. Importantly, during the consultation period before the last TG203 update (2019), OECD member countries expressed a concern over the high level of subjectivity in recording clinical observations in fish. There is also a distinct difference in the protection goals for mammalian toxicology and ecotoxicology - for mammalian assessments the protection goal is the individual whereas for ecotoxicology the protection goal is most often the population.

There was similar reluctance before acceptance of evident toxicity for certain routes of administration in mammals (despite this term being used in acute oral toxicity as early as 2002). For the inhalation route (TG433), substantive retrospective evidence was collected to support its use and provide guidance on what constitutes evident toxicity. This involved analysis of a large database of clinical signs recorded during acute inhalation studies to determine the signs which, if observed at a lower concentration, would predict death or severe toxicity in animals exposed to a higher concentration. Information on the predictivity of a wider range of signs was included with the TG as supplementary material to help guide decisions. Comparative analysis with methods which used death as an endpoint, formed the evidence that this refinement provided equivalent protection of human health, without leading to overly conservative risk assessments or classifications (Sewell et al., 2015, 2018). More recently, for human health, there is a desire to eliminate acute toxicity testing altogether. e.g. US Environmental Protection Agency (US EPA) announcement to stop animal testing for human health protection by 2035 (US EPA, 2019), and a NC3Rs/NICEATM/JRC workshop in 2017 "Towards elimination of the acute toxicity six-pack" (Prior et al., 2019).

A multi-sectoral, international, stakeholder group met in London, UK in early March 2020 to discuss what further actions were needed to improve fish welfare in TG203. The workshop was sponsored by the Department for Environment, Food and Rural Affairs (Defra) and co-organised by the Centre for Environment, Fisheries and Aquaculture Science (Cefas) and the National Centre for the Replacement, Refinement & Reduction of Animals in Research (NC3Rs). The aim of the workshop was to identify knowledge gaps impeding refinement of TG203 and identify a roadmap to address animal welfare concerns whilst maintaining environmental protection. The following objectives

were deemed prerequisites for achieving these aims.

1. To share feedback from contract research organisations (CROs) on the TG203 2019 update following a Cefas-led survey (Supplementary file 1).
2. To discuss and identify what training materials are required to standardise reporting of sublethal clinical signs of toxicity internationally.
3. To identify a viable mechanism by which data from TG203 tests conducted internationally can be collected and stored to support analysis of linkages between sublethal clinical signs, moribundity and death, and chemical mode of action.

2. The current status of fish acute toxicity testing

2.1. The origin of TG203

The notion that altered water chemistry can be hazardous to the physiology of fishes was reported over a century ago (Ringer, 1884), and the first proper reports of toxicity experiments followed in the 1920s and 1930s (Belding, 1927; Harukawa, 1923; Jones, 1938). However, it was not until the 1950s that work began to develop a standard laboratory test method for assessing industrial wastewaters for fish toxicity (Doudoroff et al., 1951; Turnbull et al., 1954) and individual new chemicals (Freeman, 1953). The methodology for acute testing was further developed in the 1960s (Alabaster and Abram, 1964; Sprague, 1969). These early studies addressed basic issues such as: the aeration of the water in relation to oxygen requirements and ventilation rates in the fish, the design of apparatus for dosing, test duration, and how to report median survival time. Data sets were also generated with known poisons such as cyanide (Alabaster and Abram, 1964), and there were even some attempts at mixtures toxicity (Brown, 1968). Methodology and use of data from the toxicity tests evolved further in the 1970s (Sprague, 1970); and by 1980 the American Society for Testing and Materials (ASTM) had released a standardised method [see review on the history of fish toxicity testing in the USA, (Hunn, 1989; ASTM, 1980)].

The OECD introduced the TG203 in 1981, and following the European Inland Fisheries Advisory Commission (EIFAC) opinion on standardising methods for fish toxicity testing (EIFAC, 1983), the OECD TG203 was updated in 1984 and 1992. The European Commission published the standard method for the acute toxicity testing of fish and other organisms in December 1992 (EEC, 1992). These early OECD versions of TG203 described how to conduct a test with juvenile fish and set the need to report the lethality estimate.

2.2. The current regulatory context

This test guideline has long served the regulation of chemicals due to its intrinsic link with lethality. Regulators around the world have been using the resulting LC_{50} values as reference points for acute toxicity over four decades, resulting in lethality data from TG203 becoming core information in global chemical management and a hard-wired concept in risk assessment. Although lethality tests are more ethically acceptable in tests with non-sentient (plant and many invertebrate) species, acute fish toxicity data are required by most, if not all, regulatory regimes globally (Burden et al., 2020). Consequently, when laboratories contracted to conduct the test were surveyed between 2014 and 2017, TG203 was by far the most commonly used vertebrate ecotoxicology study (Burden et al., 2017). The use of the acute toxicity fish data along with requirements varies depending on the regulatory scheme; these were recently discussed elsewhere (Burden et al., 2020) and as such are not repeated here. Briefly, the data generated by the test generally has three main regulatory applications:

a) Hazard identification (classification and labelling of chemicals), along with invertebrate and algae toxicity data.

---

[1] The term endpoint has different meaning in the context of toxicology and animal experimentation (see Ellis and Katsiadaki, 2020), hence we attempt a quick clarification of the terminology used thereafter. Humane endpoints refer to animals removed at a pre-defined state based upon clinical signs to limit suffering, whilst an endpoint in regulatory toxicology refers to a response variable, used to analyse treatment effects once the experiment has ended. Clinical signs, observations and sub-lethal signs are used interchangeably as they are terms employed in various test guidelines.

b) Risk assessment for a substance or discharge as fish represent a key surrogate for many aquatic species and are environmentally relevant.

c) Water quality assessment of surface waters and effluents.

However, it should be noted that many legislations governing chemical safety assessment, particularly in Europe, demand that vertebrate animal tests are conducted as a last resort and that non-animal methods are used where possible and existing data shared (e.g. article 39 of (EC regulation, 2009)), hence, vertebrate tests need to be fully justified. Examples of non-animal methods considered include QSAR predictive approaches and in the case of plant protection products determining whether toxicity can be predicted based on active substance data. The data requirements (Commission Regulations 283/2013 and 284/2013) are such that vertebrate testing may still be required (EU, 2013a, 2013b). If this is the case in order to minimise fish testing, a threshold approach to acute toxicity testing on fish should be considered (see framework in Creton et al., 2014).

For an environmental assessment under the REACH regulations, alternatives such as predictive models, read-across data or an early life stage test can be used to fulfil the acute toxicity to fish on a case by case basis (ECHA, 2021). Nevertheless, REACH standard requirements often include acute, chronic or early life stage fish toxicity tests (EC regulation, 2006). However, since REACH also enables a weight of evidence approach, then more emphasis could be placed on invertebrate data instead.

Furthermore, for establishing water quality criteria for the environment, there is no requirement to use fish mortality data. For example, the US EPA guidance on water quality criteria indicates that 96 h EC_{50}[2] values of the percentage of immobilised fish (i.e., moribundity not lethality) should be used instead of LC_{50} values for fish (US EPA et al., 1985). In environmental risk assessment, species sensitivity distributions (SSDs) can be used to derive environmental concentrations that protect 95% of the species [e.g., the HC_5, the median hazardous concentration for 5% of the organisms (Stubblefield et al., 2020)]. Given that invertebrate species are usually the more sensitive species in such distributions to predict the HC_5 value (Weyers et al., 2000; Hutchinson et al., 2003; Jeram et al., 2005), it should be questioned whether fish-based SSDs are an appropriate approach in all cases. Overall, there is certainly a case to be made to question the utility of TG203, or other fish tests that use mortality as the endpoint, in environmental protection. In addition, the variability present in species extrapolation, from acute to chronic, from freshwater to marine, etc., that are used in regulatory schemes may be larger than any experimental uncertainty in using non-lethal endpoints in fish tests.

2.3. Is TG203 fit for purpose?

TG203 is currently considered fit-for-purpose in most frameworks due to an historic understanding or acceptance that the protection goals are met through the use of the LC_{50} value generated in an appropriate assessment scheme. We consider the main barriers to alteration or substitution of the mortality endpoint revolve around the specific understanding of the protection goals, which are only just beginning to be critically examined (e.g. EFSA (European Food Safety Authority), 2016; Brown et al., 2017), and the level of protection offered by the endpoints and assessment schemes utilised.

The environmental protection goal with reference to fish acute toxicity is usually to avoid visible mortality; however, in most EU

legislation this is not explicitly stated, and the goals are in fact to avoid unacceptable effects on non-target species or similar (Brown et al., 2017). It can be assumed that most jurisdictions would consider visible mortality as an unacceptable effect from environmental exposure to a chemical. While the LC50 is qualitatively relevant to a protection goal of ensuring no unacceptable mortality (the biological endpoint being the same), quantitatively there is a need to extrapolate down to the acceptable level, which is typically managed by application of an appropriate assessment factor. It is therefore, also assumed that the assessment factor will ensure this protection goal is met. While databases of wildlife poisoning incidents do exist in various jurisdictions, which could be used to retrospectively evaluate the level of protection, the authors are not aware of any formalised attempt to do so. It should also be considered whether the sub-lethal effects observed in a highly controlled laboratory test in model species (often captive bred strains) are reflective of that seen in wild fish populations. The fitness of the test, in terms of wider environmental protection goals, is further questioned via the concept of 'ecological death' (Scott and Sloman, 2004; Rand, 1995).

It should be noted that early euthanasia of moribund fish is already regularly practiced in Europe and Canada (with EU Directive, 2010/63/EU[3] stating that, where possible death should be "substituted by more humane endpoints using clinical signs that determine the impending death"). The risk of OECD member countries adopting different approaches may compromise the mutual acceptance of data (MAD) requirements.

In summary, TG203 causes severe suffering and death, yet its translatability to real-life field situations is open to question. A shift in the regulatory paradigm is needed before the test is replaced, which we anticipate will take some time. Therefore, immediate, international, and evidence-based action is needed to achieve refinement. The discussions held during the 2-day workshop in London, identified two main refinement opportunities which are summarised in Section 3.

2.4. 3Rs opportunities

European legislation regulating the use of sentient animals for scientific purposes requires equivalent assessment and reporting of suffering for all vertebrates, regardless of taxon (EU, 2010; UK Home Office, 2014). However, the additional provisions on the use of certain animals including non-human primates suggests a wider perception that 'some animals are more equal than others'. This is also implicit in the requirement to select 'species with the lowest capacity to experience pain, suffering, distress or lasting harm' (EU, 2010), although given the difficulties in determining animal sentience[4] (Dawkins, 2012) it is no longer clear what 'lowest capacity' means, or what criteria ought to be used for decision-making. It is often still assumed that fishes are 'lower' animals, although these assumptions do not stand up to scrutiny (Brown, 2015; Message and Greenhough, 2019). Fishes used in research and testing have traditionally received less attention than terrestrial animals (Hawkins et al., 2011) as demonstrated by the relatively sparse guidance on standards of housing, husbandry and care in research establishments (EU, 2010) and greater acceptance of death in captivity, justified in part because wild fishes experience high mortality rates during early life stages (Hawkins et al., 2011). There is also a general lack of public concern about fish welfare, due to the evolutionary distance from fish, reduced empathy and the cultural acceptance of suffering during commercial and sport fishing. However, these views are changing with

---

[2] EC_{50} (the half maximal effective concentration) refers to the concentration of a drug, antibody or toxicant which induces a response halfway between the baseline and maximum after a specified exposure time. More simply, EC_{50} can be defined as the concentration required to obtain a 50% effect. The effect can be defined in contract to the LC_{50}, where the effect is death.

[3] https://ec.europa.eu/environment/chemicals/lab_animals/legislation_en.htm.

[4] Sentience can be defined as the ability to experience 'feelings that matter', such as fear, pain and pleasure. Attributes of a sentient animal include some kind of centralised nervous system, and behaviours that show they are actively avoiding painful or distressing situations and seeking out positive experiences.

increasing recognition that fishes can feel pain, experience psychological distress and have positive welfare (Balcombe, 2016).

The authors of this report believe that it is essential to ensure that fishes used in research and testing receive the same consideration as other vertebrates, and that mortality as an endpoint – and all avoidable suffering – should be critically scrutinised and every opportunity to apply the 3Rs implemented. There is a growing desire to apply the 3Rs to vertebrate ecotoxicity testing, whilst at the same time ensuring that the intended environmental protection goals are met. It is particularly important that this is considered for acute toxicity tests, which by nature cause severe suffering to test animals. While TG203 and similar standard *in vivo* tests often remain the principle or only acceptable means of generating fish acute toxicity data for new chemicals, under current regulations, it is imperative that the 3Rs are implemented as widely as possible (Burden et al., 2020).

Replacement - The ultimate aim would be for fish acute toxicity data to be generated using alternative technologies and approaches which replace the TG203 test. This includes computational approaches such as quantitative structure activity relationship (QSAR) models, which show promise for use in a regulatory context for certain purposes (e.g. Benfenati et al., 2011; Chaudhry et al., 2010; Burden et al., 2016). Investment has also been made in developing *in vitro* assays with cytotoxicity endpoints, including the rainbow trout RTgill-W1 cell line assay (ISO, 2019 and OECD – TG development in progress). Fish embryo assays have so far been the most promising potential alternative for predicting fish acute toxicity, as a (partial) replacement for fish at later life stages. While technically being *in vivo*, these tests use fish at an early life stage that is not protected under many legislations (e.g. EU, 2010), because they are not considered to have developed sentience (EFSA (European Food Safety Authority), 2005). Despite validation of the Fish Embryo Toxicity Test as an OECD test guideline (OECD, 2013) and evidence that the endpoints correlate well with those from TG203 tests (Belanger et al., 2013), the broad applicability of this assay has been questioned (ECHA (European Chemicals Agency), 2016). As a result, FET data are not currently widely accepted for regulatory purposes and work is currently underway to determine how - rather than being used as a 1:1 direct replacement of TG203 data - they can be incorporated into integrated approaches to testing and assessment (IATAs; Project 2.54: Guidance Document on IATA for Fish Acute Toxicity Testing; Paparella et al., 2021) or used in weight-of-evidence approaches (CEFIC, 2020).

Reduction – Although it has been shown that reduction to six fish per test concentration would yield LC_{50} estimates of quality similar to that obtained using the seven fish presently required, unless the slope of the concentration-response curve is low (Rufli and Springer, 2011), this was not unanimously supported by VMG-Eco. The minimum group size (n = 7 fish) stated in TG203 was reduced in the 1992 revision from 10 fish, but other ways to reduce the number of animals tested have been introduced. Importantly, the limit test and the newly incorporated threshold approaches have, and will, further reduce the number of fish used. New to the 2019 revision of TG203, was the option to omit the dilution water control when a solvent is used, reducing the number of fish by seven per test. The OECD also recently updated Guidance Document (GD) 23 on Aqueous-Phase Aquatic Toxicity Testing of Difficult Test Chemicals (OECD, 2019a, 2019b), including revisions to reduce occasions when solvents need to be used for poorly water-soluble substances. This is also intended to reduce additional solvent control groups.

Refinement - Anticipating that it will be some time before alternative approaches are routinely or widely accepted to replace TG203 data, and given that no further reduction in fish numbers (group size and treatment groups) is achievable, participants agreed that the biggest impact on animal welfare in the coming years would come from refinement approaches.

3. Refinement opportunities for TG203

The key refinement opportunities identified to limit the level and duration of suffering were:

a) Applying early endpoints sufficiently predictive of death and/or toxicity.
b) Reducing the duration of the test (< 96 h).

Both approaches would reduce suffering and could be applied alone or in combination. However, they both require collection of evidence for progress towards international adoption.

3.1. Evidence base needs on moving away from mortality as an endpoint

The most efficient way of linking clinical signs to mortality in a test where mortality is the endpoint, is to perform the test and record clinical signs as required (including moribundity) and continue until mortality is reached at the end of the test. This information is highly compatible with what is legally required in TG203 (2019) and the data could be collated for analysis. An additional benefit of such a collated dataset is that it would be a resource for validation of adverse outcome pathways relevant to basal toxicity, and thereby provide a solid foundation for predictive toxicology, a key scientific, regulatory, and societal aim. This would allow exploring the relationship between the physicochemical properties of the test substance and/or any information on mode of action, with the clinical signs and mortality data. Fish tests however present unique challenges in this context, that preclude or reduce our ability to collect certain type of information at individual level, which are summarised below.

1. Fish are maintained in populations which, compounded by fish movements, makes identification and tracking the behaviour and state of individuals over time difficult (Midtlyng et al., 2011; Rufli, 2012). Mammals and birds are typically housed in small groups, where individual identification is possible. An additional issue impeding tank observation can be poor visibility (Dennison and Ryder, 2009) as some chemicals can cause turbidity. Since the issue of individual fish identification has been a key impediment in linking clinical presentation to an individual, further details on the disadvantages of marking and future perspectives are discussed in Supplementary file 2.

2. Fish are not amenable to handling for direct clinical examination as recommended for mammals (CCAC Canadian Council on Animal Care, 1998). Handling fish out of water is considered to cause a maximal acute stress response, can damage the immuno-protective skin, and cause injury, scale loss and gill collapse (Hawkins et al., 2011; Wedemeyer, 1996). As handling could compromise the experiment (and cause additional suffering) clinical signs for fish are restricted to visual abnormalities in appearance and behaviour assessed via non-intrusive observations. Capturing fish can also prove difficult and a prolonged capture attempt may increase stress in both the individual and remaining population.

3. In mammals, physiological measurements (e.g. weight loss, body temperature changes, heart rate) can be used as clinical signs that may be predictive of mortality (CCAC Canadian Council on Animal Care, 1998). Such physiological parameters are irrelevant or not measurable in fish. Factors such as ventilation rate are expected to be more variable in fish due to changes with temperature, fish size and activity levels at the time of observation. Fish also differ to mammals in lacking vocalisation or recognisable facial expressions.

Moribundity is already being applied as an early endpoint equivalent to mortality in TG203 (OECD, 2019a, 2019b), primarily in European countries, and there was general consensus amongst workshop attendees that future inclusion in TG203 should be a refinement aim to reduce the
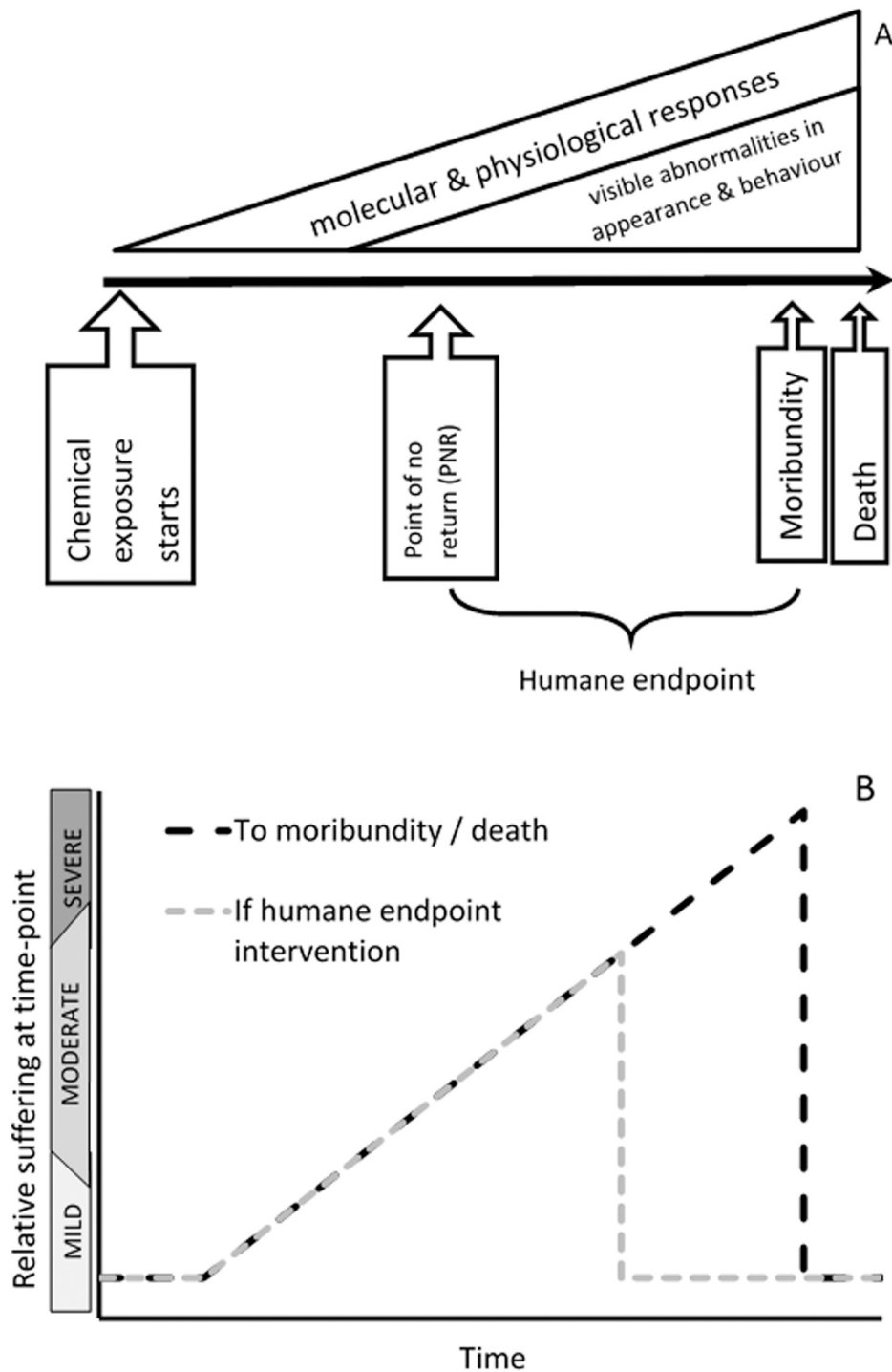
Fig. 1. Conceptual time-line of stages between initiation of exposure and death during an acute toxicity test to illustrate the application of UK CRO pre-mortality endpoints. Figure adapted from Brønstad (2019) and Wolfensohn and LLoyd (2003). See Ellis and Katsiadaki (2020) for further explanation.

level of suffering. A study has shown that the LC_{50} can be affected if moribundity, rather than mortality, is used as the endpoint. The effect of applying five different definitions of moribundity on the LC_{50}, based on the four general observations/signs required in TG203 (1992) from 512 studies, revealed that in 36–52% of the studies, the MC_{50} (concentration at which 50% of the fish were moribund) was lower than the conventional LC_{50} depending on the definition of moribund (Rufli, 2012). The study also concluded that the inclusion of the moribund criterion in TG203 would reduce the period of suffering, lowering the LC_{50} by a

factor of approximately 2 (median), and maximally by 16 (Rufli, 2012). When discussions on the TG203 update re-started at VMG-Eco in 2014, it was suggested that to produce consistent TG203 data worldwide, the following information was required (Rufli, 2012).

- A unique definition for the moribund state in fish (could be species-specific)
- Specifications on an unambiguous type of visible abnormality to be reported (quality of effect)

- Specifications on the degree of the effects to be reported (quantity of effect)

Therefore, initial efforts towards eliminating mortality focused on defining moribundity. TG203 required that 'Records are kept of visible abnormalities (e.g. loss of equilibrium, swimming behaviour, respiratory function, pigmentation, etc.)'. As such, the definition for moribundity was initially based on these four clinical signs. However, UK-based CROs that have been applying the moribundity endpoint for several years suggested that the following two clinical signs could be considered as strongly indicative of imminent death:

- Fish show little or no sign of locomotory movement often, with very slow respiratory action (immobile fish may be present at any level in the water column-not just at the bottom).
- Fish elicit little or no response to stimulation.

This description is very close to the definition of mortality in TG203: "No sign of physical movement together with no response to physical stimulation". However, no consensus was reached at VMG-Eco for adopting the UK-proposed definition. The only agreed definition included the two existing signs used for mortality, immobility and inactivity after stimulus. In summary, the lack of implementation of moribundity in place of mortality in TG203 has been due to two main technical difficulties: the lack of consensus on the definition of moribundity; and the difference between the derived toxic concentrations (MC_{50} and the LC_{50}) that regulators are familiar with and have used routinely for four decades.

Some authors debate the value of this refinement even if a definition of moribundity in fish was agreed (e.g. based on immobility and inactivity after stimulus); by that time the fish could be considered unconscious and beyond suffering (Ellis and Katsiadaki, 2020). As fish are observed at discrete observation times, if a pre-mortality endpoint is applied then fish will be euthanised at any state between first showing the defined clinical sign(s) and death. This is illustrated in Fig. 1 where the CRO endpoints are imposed on a theoretical time-line of exposure to a chemical: the earlier the pre-mortality endpoint is applied, then the greater the range of states encompassed and the "noisier" the data from the test will be. However, this is the space where most of the ethical gains can be made, provided evidence could be collected in a scientifically robust manner, supported by guidance and training.

To support implementation of pre-mortality endpoints, actions need to include (1) development of guidance and training material (glossary and atlas of clinical signs) to aid consistent recording, followed by a validation exercise (2) analysis of historical and prospective data to identify appropriate endpoints and conversion factors to enable estimation of LC_{50} values, and (3) adjusting reporting template tables to ensure systematic reporting of clinical sign data.

3.1.1. Guidance and training

TG203 (2019) included a comprehensive list of morphological and behavioural abnormalities potentially displayed by fish (with definitions). However, categorisation of such clinical signs in fish remains subjective in practice and is currently not harmonised between laboratories and/or countries. Appropriate training, both in-house and via a shared manual, is the only means by which clinical sign recording can be harmonised. An overarching guidance document, including a vetted pictorial training manual is key in addressing the issue of subjectivity in observing fish appearance and behaviour. Like any training material, the guidance should be simple, clarify binary responses (absence/presence of a sign) and include competency assessment. Competency assessment would also feedback which clinical signs are most difficult to recognise. Importantly, a key success feature of the guidance would involve OECD endorsement. There are prior examples where expert knowledge was

gathered in a form of 'guidance document' or 'supporting document', e. g. a histopathology atlas for assessing endocrine disrupting chemicals (OECD, 2010b). Training materials exist in the CROs that undertake the testing, at least in the UK, due to the Good Laboratory Practice (GLP) and ASPA[5] regulations requiring evidence of competence. This is also true for many other countries; particularly where humane endpoints are commonly applied. Discussions acknowledged that image sensitivity issues may prove an obstacle, nevertheless, it was unanimously agreed that CROs would be the ideal means of collecting material for the training manual. The fact that chemicals with different modes of action cause different clinical signs (Drummond et al., 1986), means that various signs and pre-mortality endpoints need to be defined. Adding to this complexity is the large number of fish species used and recommended for the test, that almost certainly will present variations in the clinical presentation. Nevertheless, each OECD member country tends to use a limited number of fish species for the risk assessment; hence work can be divided based on experience.

Several existing initiatives, such as the 'FISHWELL project', pictorial manual on salmon (Noble et al., 2018) could contribute to the design or content of a guidance; training material specifically designed for chemical toxicity, supported by images and videos, would be more valuable in achieving refinement in the short term. Finally, the potential need for an additional workshop, specific to the training and guidance needs and chaired by the OECD was discussed. The idea was supported as means of ensuring a direct dialogue between experienced CROs, regulators, clients and notifiers, which is necessary as a means of removing some of the perceived difficulties in fish welfare assessment.

3.1.2. Data analysis

Fish acute chemical toxicity data is a rich resource as TG203 was not only one of the first OECD test guidelines published, but also one that is highly used in both hazard and risk assessment of chemicals globally. As such it lends itself to both retrospective and prospective data analysis on the key information reported. Observations on fish appearance, respiration and swimming behaviour as well as mortality outcomes are, and have always been, reportable as key information. The first dataset to be mined for differences between mortality and moribundity endpoints (LC_{50} and MC_{50}) included a series of 512 TG203 tests performed from 1990 to 2001 (Rufli, 2012). A second dataset, collected during the period that preceded the recent TG203 update but not analysed due to the lack of linkage to individual fish, includes a further 111 studies from 10 laboratories; six laboratories from Europe (Harlan, Ibacon, Safepharm, Springborn Smithers, Brixham, and Eurofins) and four from the United States (Springborn Smithers, Wilbury, ABC, Wildlife International). All these data derived from tests using mortality as an endpoint. Many more data that can be used to enhance this dataset for meta-analysis exist, primarily within CROs, but also within industry as the sponsors. Even if there is no individual fish link, the tank level observations can provide important insights on the clinical presentation of toxicity in fish.

Retrospective data analysis of existing and future datasets could reveal the extent of fish recovering from a severe clinical picture, as data on both fish observations and mortality outcomes are reported for each day of the test. Severe clinical signs that are present on one day but absent the following day could be used as rough indicators of the frequency at which fish appear to recover. Data from CROs that apply moribundity as an endpoint will also be extremely useful. Prospective data analysis will greatly benefit from guidance and training material. The use of the full list of clinical sings in TG203 (OECD, 2019a, 2019b) can contribute to generating fundamental knowledge on baseline toxicity in fish, allowing the linking of molecular targets with key events and adverse outcomes (lethality). This pathway description is needed for

---

[5] ASPA: Animals (Scientific Procedures) Act 1986 https://www.legislation.gov.uk/ukpga/1986/14/contents'.

predictive toxicology to replace animal testing in the future altogether.

To further advance our ability to predict toxicity in the future, both retrospective and prospective data analysis will benefit from the additional information on the chemical, notably key physico-chemical properties such as K_{ow}, special structural features and suspected mode of action. Such information could allow potential links between the mode of action of the chemical and time to effect, sequence of clinical sign appearance and importantly frequency of recovery from a moribund state, to be explored (e.g. is the expression of transient but severe clinical signs in fish associated with some form of narcosis?).

3.1.3. Reporting templates

Harmonisation of the reporting template was one of the workshop recommendations that has already been achieved. The OECD harmonised template (OHT) for reporting short term toxicity to fish information (OECD template#41, V6.5 and associated tables, 2020a) was recently updated and includes fields to record a summary of results including information on test organism size/age (mean wet weight or length), test type (flow through, static, static renewal) and the derived LC_{50} (95% C. I.), NOEL (Probit Slope) and EC_{50} (95% C.I.) along with tables on mortality and sublethal clinical signs as presented in TG203 (2019) Annex 4 (OECD, 2019a, 2019b). Increased reporting of clinical signs in substance property or ecological databased/reporting tools in the public domain, along with more consistent recording (e.g. through use of a technical manual or the recently updated TG203) will enable a continuous retrospective assessment of clinical signs in the future.

Importantly, the use of the harmonised template including the predetermined tables should be strongly encouraged by regulators as non-endpoint data entry (e.g. data other than the LC_{50}) is often optional within individual regulatory regimes. Use could be encouraged via quality observation requests from regulators as part of compliance checks/substance evaluations and supported via a targeted working group within the industry operating via sponsors and CROs.

3.2. Evidence based needs for reducing the test duration

Decreasing suffering can not only come from addressing the level of suffering but also its duration. The original reason behind the 96 h duration for fish toxicity testing, and eventually for TG203 was discussed during the workshop. Participants were unaware of multiple lines of robust scientific reasons, with the exception of a general agreement that most chemicals reach a steady state within that period as reported by McCarty (2012). Nevertheless, even this analysis that was based on 777 tests with 644 chemicals conducted by a standard 96 h LC_{50} testing protocol on the fathead minnow Russom et al. (1997) concluded that a steady-state LC_{50} is not consistently met in test results obtained following standard 96 h LC_{50} protocols (McCarty, 2012). Indeed, scientific arguments around the test substance coming into equilibrium with the internal organs of the fish over 96 h, and therefore offering a 'true' exposure, do not stand up to scrutiny of the animal physiology. In sub-lethal conditions where the gills likely remain functional, the time for steady-state equilibrium between the external water and the fish is many days, often several weeks, as shown by bioaccumulation studies (Veith et al., 1979). In acute aqueous testing, where gill pathology and respiratory distress are often the cause of death, the concept of equilibria is undermined. In such situations, the test substance may quickly diffuse through the damaged gill into the blood supply. It is not a controlled exposure with respect to dose in the internal organs. So, there is no justification for keeping the test duration as 96 h. One might argue to make the test duration 48 h or 24 h from the perspective of internal dose. For substances that are actively taken up through transport pathways in the gills, such as Na^+, it might take as little as 30 s for the substance to appear in the blood (Handy and Eddy, 2004), and for organic chemicals with uptake rate constants of the order of ml/g/h (e. g., Erickson et al., 2006), the dose will appear in the blood within a few hours. Thus, from a dosimetry perspective, TG203 could be substantially
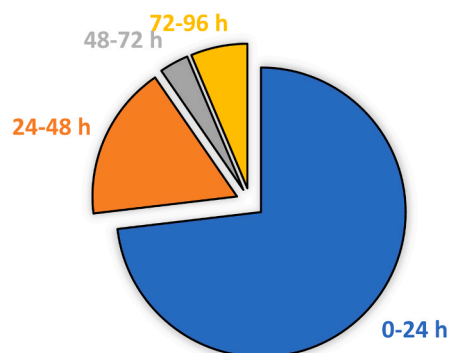

Fig. 2. Number of fish removed as dead or moribund per 24 h period (Total = 521, 39 TG203 tests from 2 UK-based CROs).

shorter than 96 h. Workshop participants agreed that reducing the test's duration should be considered, as a refinement opportunity. It was also noted that the acute toxicity test using *Daphnia* is 48 h long.

This idea gained further support as CRO interviews revealed that most, if not all signs of toxicity are evident within the first 24 h, at least in high concentrations, and that mortalities become fewer over time. Pilot data, readily available from two CRO's, were analysed shortly after the workshop. The sample size is small but indicative of the point that acute toxicity signs tend to appear very early after the onset of exposure (Fig. 2).

Furthermore, reducing test duration would have additional benefits: from a failure perspective as the risk of non-procedural mortalities in control fish would decrease; and from a welfare (hunger) perspective, as fish are fasted for 1–2 days before, and for the duration of the current test. Although there is still little scientific basis on maximum periods of feed withdrawal in fish, these never exceed 72 h in most codes and standards following recommendations of the Farm Animal Welfare Committee's (Defra) report (Farm Animal Welfare Committee (FAWC), 1996). Although it is generally accepted that wild fish can withstand longer periods without feeding than warm-blooded animals, farmed and experimental fish may have become habituated in regular feeding and as such food withdrawal may be stressful. In addition, sudden feed withdrawal may reduce welfare because aggression may increase (Farm Animal Welfare Committee (FAWC), 2014).

Importantly, the calculation of LC_{50} at different time points (24 h, 48 h, 72 h, 96 h) using both existing and newly collected datasets should be a straightforward process. It was thought that many CROs, who are the custodians of the data for 5 years under GLP regulations, would be willing to share them provided they remain anonymous. Sponsors may also have a strong interest in data sharing, and where intellectual property issues allow, facilitate the provision of additional test substance information (i.e. chemical structure, K_{ow} and mode of action where known).

Given that some fish, even if relatively few, will die on days 3 and 4 of exposure, it was safely assumed that this refinement would generally result in higher (less protective) LC_{50} values and as such, may not be acceptable to regulators. Risk and hazard assessments are often binary (e.g. pass/fail or category a or b) hence even modest changes in the derived LC_{50} can have major implications for risk, potentially adding the need for revised assessment factors. However, the participants agreed a working group could define the magnitude and quality of evidence needed to allow efficient data mining for correction factors.

Acknowledging the importance of the current LC_{50} benchmark, to accommodate difficulties with risk assessment, the effect of both approaches (application of early endpoints and reduction of test duration), on LC_{50} should be investigated, singly and in combination. Applying
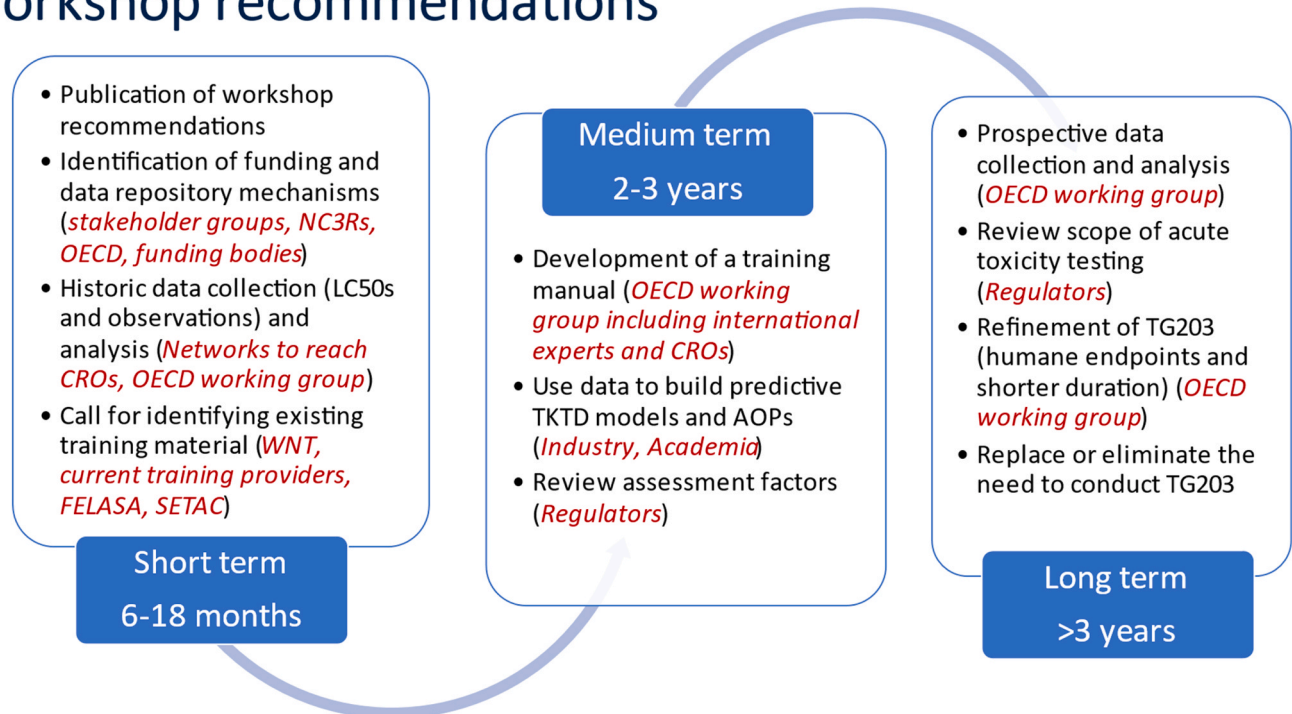
Workshop recommendations

- Publication of workshop recommendations
- Identification of funding and data repository mechanisms (*stakeholder groups, NC3Rs, OECD, funding bodies*)
- Historic data collection (LC50s and observations) and analysis (*Networks to reach CROs, OECD working group*)
- Call for identifying existing training material (*WNT, current training providers, FELASA, SETAC*)

Short term
6-18 months

Medium term
2-3 years

- Development of a training manual (*OECD working group including international experts and CROs*)
- Use data to build predictive TKTD models and AOPs (*Industry, Academia*)
- Review assessment factors (*Regulators*)

- Prospective data collection and analysis (*OECD working group*)
- Review scope of acute toxicity testing (*Regulators*)
- Refinement of TG203 (humane endpoints and shorter duration) (*OECD working group*)
- Replace or eliminate the need to conduct TG203

Long term
>3 years

Fig. 3. Recommendations on the roadmap to refine TG203.

both refinements that push the LC_{50} in different directions, may just achieve a negligible drift from current risk assessment, whilst vastly improving fish welfare and building solid toxicological knowledge.

Furthermore these recommendations to characterise sublethal effects, can also be taken forward for other regulatory acute fish toxicity tests in addition to and in alignment with TG203, namely the International Standards Organisation (ISO) acute fish toxicity methods (ISO 7346-1_1996 static and ISO 7346-2_1996 semi-static) as part of the ISO 5 year systematic review.

In summary, a highly fertile ground of discussions took place around considering reducing the exposure period for estimating fish acute toxicity, not only as means of reducing animal suffering but also as means of maximising toxicological information derived from animal studies.

4. Workshop recommendations

During the last part of the workshop discussions, participants summarised the refinement opportunities for TG203 in a realistic roadmap, highlighting the key issues that need to be addressed in order to harness these refinement opportunities in coming years.

The current TG203 requires the recording of observations that fish may display, with the option to record a comprehensive list of clinical signs, in addition to mortality outcomes. Although this is common practice in countries such as the UK, harmonisation of training in clinical sign recognition, interpretation and reporting needs to be addressed at a global level to support consistency and quality of acute toxicity data. Without this clear guidance, it can be safely assumed that individual differences at either CRO or country level will be substantial, not only in how signs are identified but the terminologies used, and how they may be grouped together. This is critical to enable a data-driven evaluation of clinical observations that could potentially be useful in defining evident toxicity, including its relationship with lethality and recovery.

Although several training manuals and online resources exist for fish health assessment, notably as a result of the increasing importance of farmed fish as a food source, but also as laboratory models (e.g.

zebrafish), discussions concluded that the guidance should be focused on the clinical presentation and progression of chemical toxicity addressing both toxicological and welfare interests. Countries where application of humane endpoints during scientific procedures on fish occurs (e.g. UK), should take the lead in collecting relevant material and road test its fitness. However, since this guidance should cover all fish species used in the acute toxicity testing and different countries prefer certain species, an international effort is needed for its completion. For this reason, the participants agreed that the process would be best initiated via a project proposal to the OECD, on the basis of this workshop report, such that the OECD could request member countries to provide input where available and the formation of an OECD working group that will overview the guidance document. It was acknowledged however that both the training material and the guidance document will take time to complete via this process, with optimistic scenarios spanning three years.

In the meantime, mining historic and current TG203 data in the form of predetermined tables, as required for reporting, is key in calculating how frequently recovery or lack of deterioration from severe clinical signs, that otherwise would progress to lethality, occurs during fish acute toxicity testing. The summary tables, theoretically required for reporting, are perfectly suitable for extracting this information, even if the granularity in the description of clinical picture is low. The frequency of table usage by CROs when reporting the studies was questioned, however. A quick and easy 'fix' can come from regulatory bodies such as the European Food Safety Authority (EFSA), the European Chemicals Agency (ECHA) and the US EPA recommending their use as good practice and highlighting omissions in quality or compliance checks. The data needed to support refinement in mammalian toxicity testing were collected via use of professional networks. A similar exercise could be conducted for fish, even before funding is secured.

Furthermore, reducing the test duration was also identified as a potentially 'low hanging fruit' refinement opportunity as data on LC_{50} progression on 24, 48, 72 and 96 h can easily be pooled. Professional networks of stakeholder groups can request the data in parallel with the clinical signs.

One of the most difficult workshop tasks was the identification of organisation(s) that can act as the data repository that provides quality assurance and curation (not necessarily analysis). The NC3Rs for example has both interest, capability and experience, via the mammalian data exercise (Sewell, 2015, 2018). Another option could be an OECD-managed web portal for data repository, but this would first require provision of funds.

Importantly, it was agreed that alongside the guidance and training resource and the collection and analysis of data-driven evidence, a shift in regulatory paradigm may also be required, at least in the way assessment factors are applied. Environmental protection levels are not necessarily synonymous to acute lethality estimates via the LC_{50}. Besides, even transient severe suffering may translate into death in the real environment (high probability of predation), presenting no clear justification for the prolonged suffering experienced by test fish. Equally, the translation of sub-lethal effects in response to toxicants observed in fish in the laboratory has not been demonstrated to those relevant fish in the environment. The systematic evaluation of clinical signs as predictors of toxicity and lethality, will undoubtedly result in the identification of a group of signs, that are reasonably predictive of acute toxicity and lethality. In addition, the scope of a fish acute toxicity test should be in parallel evaluated as if reducing the exposure duration proves to have a negligible effect on LC_{50}, the welfare benefit could be substantial.

Key to the acceleration of the science needed to modernise current toxicological approaches that rely on animal data, is to make the data publicly available for further scrutiny and analysis. Several initiatives to this effect exist as imminent plans, for example the OECD is creating a Global Chemicals Knowledge Base and due to the recent Transparency Regulation (EU, 2019),[6] EFSA is developing the Metapath format, an international database on pesticide metabolism (Kolanczyk et al., 2013; OECD, 2020b) allowing publication of the non-confidential content of pesticide dossiers in a searchable electronic format. This level of transparency will enable the analysis of the large quantity of regulatory data available. Sponsors can and already play a significant role by adopting transparency policies and data sharing. Finally, a similar initiative can aim at strongly encouraging reporting of academic research in the same format (Fig. 3).

CRediT authorship contribution statement

All authors contributed to discussions during the workshop and writing several parts of the manuscript, both the original and further drafts, reviews, and editing. The lead author, IK was responsible in assimilating the views expressed into a coherent manuscript.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

---

[6] The Transparency Regulation, which amends the General Food Law (GFL), means that EFSA has new requirements for capturing, managing, handling and distributing data on plant protection products (PPP). These changes require the specification of data formats for regulated product dossiers and allow documents to be submitted, searched, copied and printed, while ensuring compliance with legal requirements. It has been decided to use IUCLID formats and the IUCLID tool (managed by the European Chemicals Agency – ECHA) for data preparation, electronic submission and management of pesticide dossiers, by means of the ECHA Cloud platform. Furthermore, it has been decided that applicants should provide data on metabolism in the areas of residues and mammalian toxicology as attachments generated with the composer software of the MetaPath software package. https://ec.europa.eu/food/safety/general_food_law/implementation-transparency-regulation_en.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.ecoenv.2021.112585.

References

Alabaster, J.S., Abram, F.S.H., 1964. Development and use of a direct method for evaluating toxicity in fish. In: Advances in Water Pollution Research. Pergamon Press, Oxford, pp. 41–54.

ASTM, 1980. Standard Practice for Conducting Acute Toxicity Tests with Fishes, Macroinvertebrates, and Amphibians, vol. E729-7.

Balcombe, J., 2016. What a Fish Knows: The Inner Lives of Our Underwater Cousins. Farrar, Straus & Giroux, New York.

Belanger, S.E., Rawlings, J.M., 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.

Belding, D.L., 1927. Toxicity experiments with fish in reference to trade waste pollution. Trans. Am. Fish. Soc. 57, 100–119.

Benfenati, E., Toropov, A.A., Toropova, A.P., Manganaro, A., Gonella Diaza, R., 2011. CORAL Software: QSAR for anticancer agents. Chem. Biol. Drug Des. 77, 471–476. https://doi.org/10.1111/j.1747-0285.2011.01117.x.

Brønstad, A., 2019. Severity and humane endpoints in fish research. In: Severity and Humane Endpoints in Fish Research. Fondazione Guido Bernardini.

Brown, A.R., Whale, G., Jackson, M., Marshall, S., Hamer, M., Solga, A., Kabouw, P., Galay-Burgos, M., Woods, R., Nadzialek, S., Maltby, L., 2017. Toward the definition of specific protection goals for the environmental risk assessment of chemicals: a perspective on environmental regulation in Europe. Integr. Environ. Assess. Manag. 13 (1), 17–37.

Brown, C., 2015. Fish intelligence, sentience, and ethics. Anim. Cogn. 18, 1–17.

Brown, V.M., 1968. The calculation of the acute toxicity of mixtures of poisons to rainbow trout. Water Res. 2, 723–733.

Burden, N., Maynard, S.K., Weltje, L., Wheeler, J.R., 2016. The utility of QSARs in predicting acute fish toxicity of pesticide metabolites: a retrospective validation approach. Reg. Toxicol. Pharmacol. 80, 241–246.

Burden, N., Gellatly, N., Benstead, R., Benyon, K., Blickley, T.M., Clook, M., Doyle, I., Edwards, P., Handley, J., Katsiadaki, I., Lillicrap, A., Mead, C., Ryder, K., Salinas, E., Wheeler, J., Hutchinson, T.H., 2017. Reducing repetition of regulatory vertebrate ecotoxicology studies. Integr. Environ. Assess. Manag. 13, 955–957.

Burden, N., Benstead, R., Benyon, K., Clook, M., Green, C., Handley, J., Harper, N., Maynard, S.K., Mead, C., Pearson, A., Ryder, K., Sheahan, D., van Egmond, R., Wheeler, J.R., Hutchinson, T.H., 2020. Key opportunities to replace, reduce, and refine regulatory fish acute toxicity tests. Environ. Toxicol. Chem. 39 (10), 2076–2089.

CCAC (Canadian Council on Animal Care), 1998. CCAC Guidelines On: Choosing an appropriate endpoint in experiments using animals for research, teaching and testing, 24 p. [Online]. ⟨http://www.ccac.ca/en/CCAC_Programs/Guidelines_Policies/PDFs/APPOPEN.pdf⟩.

CEFIC, 2020. LRI-ECO51: Integrating the fish embryo test into the weight of evidence to inform acute fish toxicity. ⟨http://cefic-lri.org/request-for-proposals/lri-eco51-integrating-the-fish-embryo-test-into-the-weight-of-evidence-to-inform-acute-fish-toxicity/⟩.

Chaudhry, Q., Piclin, N., Cotterill, J., Cotterill, J., Pintore, M., Price, N.R., Chrétien, J.R., Roncaglioni, A., 2010. Global QSAR models of skin sensitisers for regulatory purposes. Chem. Cent. J. 4 Suppl 1, 5. https://doi.org/10.1186/1752-153X-4-S1-S5.

Creton, S., Clook, M., Wheeler, J.R., 2014. Application of the threshold approach for acute fish toxicity testing to plant protection products: a proposed framework. Chemosphere 96, 195–200.

Dawkins, M.S., 2012. Why Animals Matter. Animal Consciousness, Animal Welfare and Human Well-being. Oxford University Press, p. 209.

Dennison, N., Ryder, K., 2009. The challenges of using humane endpoints in fish research. ⟨https://norecopa.no/media/6272/abstract-ryder-endpoints.pdf⟩. (Accessed September 2020).

Doudoroff, P., Anderson, B.G., Burdick, G.E., Galtsoff, P.S., Hart, W.B., Patrick, R., Strong, E.R., Surber, W., Van Horn, W.M., 1951. Bio-assay methods for the evaluation of acute toxicity of industrial wastes to fish. Sew. Ind. Waste 23, 1380–1397.

Drummond, R.A., Russom, C.L., Geiger, D.L., DeFoe, D.L., 1986. Behavioral and morphological changes in fathead minnow (Pimephales promelas) as diagnostic endpoints for screening chemicals according to mode of action. Aquat. Toxicol. Environ. Fate 9, 415–435.

EC regulation No 1907/2006 of the European Parliament and of the Council of 18 December 2006 concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH), establishing a European Chemicals Agency, amending Directive 1999/45/EC and repealing Council Regulation (EEC) No 793/93 and Commission Regulation (EC) No 1488/94 as well as Council Directive 76/769/

EEC and Commission Directives 91/155/EEC, 93/67/EEC, 93/105/EC and 2000/21/EC. Official Journal of the European Union: L 396.

EC regulation No 1107/2009 (article 39) of the European Parliament and of the council of 21 October 2009 concerning the placing of plant protection products on the market and repealing Council Directives 79/117/EEC and 91/414/EEC.

ECHA (European Chemicals Agency), 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. UFZ-ECHAreport 2014-341 (europa.eu).

Edwards, P.J., Leopold, A., Beavers, J.B., Springer, T.A., Chapman, P., Maynard, S.K., Hubbard, P., 2017. More for less: analysis of the performance of avian acute oral guideline OECD 223 from empirical data. Integr. Environ. Assess. Manag. 13 (5), 906–914.

EEC, 1992. Annex to Commission Directive 92/69/EEC of 31 July 1992 adapting to technical progress for the seventeenth time Council Directive 67/548/EEC on the approximation of laws, regulations and administrative provisions relating to the classification, packaging and labelling of dangerous substances. Off. J. Eur. Communities 35.

EFSA (European Food Safety Authority), 2005. Opinion of the Scientific Panel on Animal Health and Welfare on a request from the Commission related to the aspects of the biology and welfare of animals used for experimental and other scientific purposes. EFSA J. 292, 1–46.

EFSA (European Food Safety Authority), 2013. Guidance on tiered risk assessment for plant protection products for aquatic organisms in edge-of-field surface waters. EFSA J. 11 (7), 3290.

EFSA (European Food Safety Authority), 2016. Guidance to define specific protection goals for environmental risk assessment at EFSA, in relation to biodiversity and ecosystem services. EFSA J. 14 (6), e04499.

EIFAC (European Inland Fisheries Advisory Commission), 1983. Working Party on Toxicity Testing Procedures. Revised report on fish toxicity testing procedures, vol. T24, Rev.1, Rome.

Ellis, T., Katsiadaki, I., 2020. Clarification of end-points and moribundity for severe (fish) experiments. Laboratory Animals. December 2020. https://doi.org/10.1177/0023677220971002.

Erickson, R.J., McKim, J.M., Lien, G.J., Hoffman, A.D., Batterman, S.L., 2006. Uptake and elimination of ionizable organic chemicals at fish gills: II. Observed and predicted effects of pH, alkalinity, and chemical properties. Environ. Toxicol. Chem. Int. J. 25 (6), 1522–1532.

EU, 2013a. Commission Regulation (EU) 283/2013. Setting out the data requirements for active substances, in accordance with Regulation (EC) No 1107/2009 of the European Parliament and of the Council concerning the pacing of plant protection products on the market. Official Journal of the European Union: L 93/1.

EU, 2013b. Commission Regulation (EU) 284/2013. Setting out the data requirements for plant protection products, in accordance with Regulation (EC) No 1107/2009 of the European Parliament and of the Council concerning the pacing of plant protection products on the market. Official Journal of the European Union: L 93/85.

EU, 2010. Directive 2010/63/EU of the European Parliament and of the Council of 22 September 2010 on the protection of animals used for scientific purposes, Environmental Law and Management.

EU, 2019. Regulation (EU) 2019/1381 of the European Parliament and of the Council of 20 June 2019 on the transparency and sustainability of the EU risk assessment in the food chain and amending Regulations (EC) No 178/2002, (EC) No 1829/2003, (EC) No 1831/2003, (EC) No 2065/2003, (EC) No 1935/2004, (EC) No 1331/2008, (EC) No 1107/2009, (EU) 2015/2283 and Directive 2001/18/EC.

Farm Animal Welfare Committee (FAWC), 1996. Report on the welfare of farmed fish, PDF, 43 pages. ⟨https://www.gov.uk/government/publications/fawc-report-on-the-welfare-of-farmed-fish⟩.

Farm Animal Welfare Committee (FAWC), 2014. Opinion on the Welfare of Farmed Fish at the Time of Killing. PDF, 36 pages. ⟨https://www.gov.uk/government/publications/fawc-advice-on-farmed-fish-welfare⟩.

Freeman, L., 1953. A standardized method for determining toxicity of pure compounds to fish. Sew. Ind. Wastes 25, 845–848.

Handy, R.D., Eddy, F.B., 2004. Transport of solutes across biological membranes in eukaryotes: an environmental perspective. In: van Leeuwen, H.P., Köster, W. (Eds.), Physicochemical Kinetics and Transport at Chemical-Biological Interphases. IUPAC Series. John Wiley, Chichester, pp. 337–356.

Harukawa, C., 1923. Preliminary report on the toxicity of colloidal sulphur to fish. Trans. Am. Fish. Soc. 52, 219–224.

Hawkins, P., Dennison, N., Goodman, G., Hetherington, S., Llywelyn-Jones, S., Ryder, K., Smith, A.J., 2011. Guidance on the severity classification of scientific procedures involving fish: report of a Working Group appointed by the Norwegian Consensus-Platform for the Replacement, Reduction and Refinement of animal experiments (Norecopa). Lab. Anim. 45, 219–224.

Hunn, J.B., 1989. History with acute toxicity tests with fish, 1863-1987. Investig. Fish Control 98, 1–20.

Hutchinson, T.H., Barrett, S., Buzby, M., Constable, D., Hartmann, A., Hayes, E., Huggett, D., Länge, R., Lillicrap, A.D., Straub, J.O., Thompson, R.S., 2003. A strategy to reduce the numbers of fish used in acute ecotoxicity testing of pharmaceuticals. Environ. Toxicol. Chem. 22, 3031–3036.

IPBES, 2019. Global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services. In: Brondizio, E.S., Settele, J., Díaz, S., Ngo, H.T. (Eds.), IPBES Secretariat, Bonn, Germany, 56 p.

ISO, 2019. Standard 21115:2019 water quality — determination of acute toxicity of water samples and chemicals to a fish gill cell line (RTgill-W1). Geneva, Switzerland.

Jeram, S., Sintes, J.M., Halder, M., Fentanes, J.B., Sokull-Klüttgen, B., Hutchinson, T.H., 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 (2), 218–224.

Jones, J.E., 1938. The relative toxicity of salts of lead, zinc and copper to the stickleback (*Gasterosteus aculeatus* L.) and the effect of calcium on the toxicity of lead and zinc salts. J. Exp. Biol. 15, 394–407.

Kolanczyk, R.C., Schmieder, P., Jones, W.J., Mekenyan, O.G., Chapkanov, A., Temelkov, S., Kotov, S., Velikova, M., Kamenska, V., Vasilev, K., Veith, G.D., 2013. MetaPath: an electronic knowledge base for collating, exchanging and analyzing case studies of xenobiotic metabolism. Regul. Toxicol. Pharmacol. 63 (1), 84–96, 63 (1):84-96.

Maynard, S.K., Edwards, P., Wheeler, J.R., 2014. Saving two birds with one stone: using active substance avian acute toxicity data to predict formulated plant protection product toxicity. Environ. Toxicol. Chem. 33 (7), 1578–1583.

McCarty, L.S., 2012. Model validation in aquatic toxicity testing: implications for regulatory practice. Regul. Toxicol. Pharmacol. 63, 353–362.

Message, R., Greenhough, B., 2019. But it's just a fish": understanding the challenges of applying the 3Rs in laboratory aquariums in the UK. Animals 9 (12), 1075. https://doi.org/10.3390/ani9121075.

Midtlyng, P.J., Hendriksen, C., Balks, E., Bruckner, L., Elsken, L., Evensen, Ø., Fyrand, K., Guy, A., Halder, M., Hawkins, P., Kisen, G., Romstad, A.B., Salonius, K., Smith, P., Sneddon, L.U., 2011. Three Rs approaches in the production and quality control of fish vaccines. Biologicals 39, 117–128.

Mielke, H., Strickland, J., Jacobs, M.N., Mehta, J.M., 2017. Biometrical evaluation of the performance of the revised OECD Test Guideline 402 for assessing acute dermal toxicity. Regul. Toxicol. Pharmacol. 89, 26–39.

Noble, C., Gismervik, K., Iversen, M.H., Kolarevic, J., Nilsson, J., Stien, L.H., Turnbull, J. F. (Eds.), 2018. Welfare indicators for farmed Atlantic salmon: tools for assessing fish welfare 351pp. ⟨https://nofima.no/en/fishwell/⟩.

OECD, 2020a. OECD Harmonised Template #41, Short-term toxicity to fish. ⟨https://www.oecd.org/ehs/templates/OHT-41-endpoint-study-record-short-term-toxicity-to-fish-v6.5-Oct-2020.doc⟩. Accessed March 2021.

OECD, 2020b. OECD QSAR toolbox ⟨https://www.oecd.org/chemicalsafety/risk-assessment/oecd-qsar-toolbox.htm⟩. Accessed March 2021.

OECD, 2002. Guidance Document on the Recognition, Assessment and Use of Clinical Signs as Human Endpoints for Experimental Animals Used in Safety Evaluation, OECD Series on Testing and Assessment, No. 19. OECD Publishing, Paris.

OECD, 2010a. Short Guidance on the Threshold Approach for Acute Fish Toxicity. Series on Testing and Assessment No. 126. OECD, Paris.

OECD, 2010b. Guidance Document for the Diagnosis of Endocrine-Related Histopathology of Fish Gonads, OECD Series on Testing and Assessment, No. 123. OECD Publishing, Paris.

OECD, 2013. Test Guideline 236: Fish Embryo Acute Toxicity (FET) Test, OECD Guidelines for the Testing of Chemicals, Section 2. OECD Publishing, Paris.

OECD, 2014. Fish Toxicity Testing Framework, OECD Series on Testing and Assessment, No. 177. OECD Publishing, Paris.

OECD, 2015. Test Guideline 240 Medaka Extended One Generation Reproduction Test (MEOGRT). OECD Guidelines for the Testing of Chemicals, Section 2. OECD Publishing, Paris.

OECD, 2019a. Test Guideline 203. Fish, Acute Toxicity Test. OECD Guidelines for the Testing of Chemicals, Section 2. OECD Publishing, Paris.

OECD, 2019b. Guidance Document on Aquatic Toxicity Testing of Difficult Substances and Mixtures, OECD Series on Testing and Assessment. OECD Publishing, Paris.

Paparella, M., Scholz, S., Belanger, S., Braunbeck, T., Bicherel, P., Connors, K., Faßbender, C., Halder, M., Lillicrap, A., Liska, R., Schirmer, K., Stoddart, G., Thomas, P., Walter-Rohde, S., 2021. Limitations and uncertainties of acute fish toxicity assessments can be reduced using alternative methods. ALTEX 38 (1), 20–32.

Prior, H., Casey, W., Kimber, I., Whelan, M., Sewell, F., 2019. Reflections on the progress towards non-animal methods for acute toxicity testing of chemicals. Regul. Toxicol. Pharmacol. 102, 30–33.

Rand, G.M. (Ed.), 1995. Fundamentals of Aquatic Toxicology: Effects, Environmental Fate, and Risk Assessment, second ed. CRC Press, Florida.

Ringer, S., 1884. Concerning the influence of saline media on fish etc. J. Physiol. 5, 98–115.

Rufli, H., 2012. Introduction of moribund category to OECD fish acute test and its effect on suffering and LC$_{50}$ values. Environ. Toxicol. Chem. 31, 1107–1112.

Rufli, H., Springer, T.A., 2011. Can we reduce the number of fish in the OECD acute test? Environ. Toxicol. Chem. 30 (4), 1006–1011.

Russell, W.M.S., Burch, R.L., 1959. The Principles of Humane Experimental Technique. Methuen, London. ISBN 0900767782.

Russom, C.L., Bradbury, S.P., Broderius, S.J., Hammermeister, D.E., Drummond, R.A., 1997. Predicting modes of toxic action from chemical structure: acute toxicity in the fathead minnow (Pimephales promelas). Environ. Toxicol. Chem. 16, 948–967.

Scott, G.R., Sloman, K.A., 2004. The effects of environmental pollutants on complex fish behavior: integrating behavioural and physiological indicators of toxicity. Aquat. Toxicol. 68, 369–392.

Sewell, F., Ragan, I, Marczylo, T., Anderson, B., Braun, A., Casey, W., Dennison, N., Griffiths, D., Guest, R., Holmes, T., van Huygevoort, T., Indans, I., Kenny, T., Kojima, H., Lee, K., Prieto, P., Smith, P., Smedley, J., Stokes, W.S., Wnorowski, G., Horgan, G., 2015. A global initiative to refine acute inhalation studies through the use of 'evident toxicity' as an endpoint: towards adoption of the fixed concentration procedure. Regul. Toxicol. Pharmacol. 73 (3), 770–779.

Sewell, F., Ragan, I., Indans, I., Marczylo, T., Stallard, N., Griffiths, D., Holmes, T., Smith, P., Horgan, G., 2018. An evaluation of the fixed concentration procedure for assessment of acute inhalation toxicity. Regul. Toxicol. Pharmacol. 94, 22–32.

Sprague, J.B., 1969. Measurement of pollutant toxicity to fish. I. Bioassay methods for acute toxicity. Water Res. 3, 793–821.

Sprague, J.B., 1970. Measurement of pollutant toxicity to fish. II. Utilizing and applying bioassay results. Water Res. 4, 3–32.

Stubblefield, W.A., Van Genderen, E., Cardwell, A.S., Heijerick, D.G., Janssen, C.R., De Schamphelaere, K.A., 2020. Acute and chronic toxicity of cobalt to freshwater organisms: using a species sensitivity distribution approach to establish international water quality standards. Environ. Toxicol. Chem. 39 (4), 799–811.

Turnbull, H., DeMann, J.G., Weston, R.F., 1954. Toxicity of various refinery materials to fresh water fish. Ind. Eng. Chem. 46, 324–333.

UK Home Office, 2014. Guidance on the Operation of the Animals (Scientific Procedures) Act 1986, Home Office. https://doi.org/ISBN 1474100287- 9781474100281.

UNEP, 2019. Global Chemicals Outlook II ⟨https://www.unenvironment.org/explore-top ics/chemicals-waste/what-we-do/policy-and-governance/global-chemicals-outloo k⟩.

US EPA, 2019. Administrator Memo Prioritizing Efforts to Reduce Animal Testing, ⟨https://www.epa.gov/sites/production/files/2019–09/documents/image2019–09 -09–231249.pdf⟩.

US EPA, Stephen, C.E., Mount, D.I., Hansen, D.J., Gentile, J.R., Chapman, G.A., Brungs, W.A., 1985. Guidelines for Deriving Numerical National Water Quality Criteria for the Protection Of Aquatic Organisms and Their Uses. PB85-227049. ⟨https://www. epa.gov/sites/production/files/2016–02/documents/guidelines-water-quality-crite ria.pdf⟩.

Veith, G.D., DeFoe, D.L., Bergstedt, B.V., 1979. Measuring and estimating the bioconcentration factor of chemicals in fish. J. Fish. Board Can. 36 (9), 1040–1048.

Wedemeyer, G.A., 1996. Physiology of Fish in Intensive Culture Systems, first ed. Springer US.

Weyers, A., Sokull-Klüttgen, B., Baraibar-Fentanes, J., Vollmer, G., 2000. Acute toxicity data: a comprehensive comparison of results of fish, Daphnia and algae tests with new substances notified in the EU. Environ. Toxicol. Chem. 19, 1931–1933.

Wolfensohn, S., LLoyd, M., 2003. Handbook of Laboratory Animal Management and Welfare. Blackwell.