

Multi-Criteria Decision Analysis of Test Endpoints for Detecting the Effects of Endocrine Active Substances in Fish Full Life Cycle Tests

Mark Crane,^{†*} Melanie Gross,[†] Peter Matthiessen,[‡] Gerald T. Ankley,[§] Stephen Axford,^{||} Poul Bjerregaard,[#] Ross Brown,^{††} Peter Chapman,^{‡‡} Michael Dorgeloh,^{§§} Malyka Galay-Burgos,^{||||} John Green,^{##} Charles Hazlerigg,^{†††} John Janssen,^{§§§} Kai Lorenzen,^{‡‡‡} Joanne Parrott,^{|||||} Hans Ruffli,^{###} Christoph Schäfers,^{††††} Masanori Seki,^{‡‡‡‡} Hans-Christian Stolzenberg,^{§§§§} Nelly van der Hoeven,^{|||||||} Dick Vethaak,^{####} Ian J. Winfield,^{†††††} Sabine Zok,^{‡‡‡‡‡} and James Wheeler^{§§§§§}

[†]WCA Environment Limited, Brunel House, Volunteer Way, Faringdon, Oxfordshire, SN7 7YR, United Kingdom

[‡]Old School House, Brow Edge, Backbarrow, Ulverston, Cumbria LA12 8QX, United Kingdom

[§]US Environmental Protection Agency, Mid-Continent Ecology Division, Duluth, Minnesota, USA

^{||}Environment Agency of England and Wales, Coverdale House, York, YO30 4GZ, United Kingdom

[#]University of Southern Denmark, Campusvej 55, 5230 Odense, Denmark

^{††}AstraZeneca, Brixham Environmental Laboratory, Freshwater Quarry, Brixham, Devon, TQ5 8BA, United Kingdom

^{‡‡}Unilever SEAC, Colworth House, Sharnbrook, Bedford, MK44 1LQ, United Kingdom

^{§§}Bayer Crop Science, Monheim, 6620 Germany

^{||||}ECETOC, Avenue Edmond Van Nieuwenhuysse 4 Bte 6, B-1160 Brussels, Belgium

^{##}Dupont Applied Statistics Group, PO BOX 27001, Richmond, Virginia 23261, USA

^{†††}Ecotoxicology and Aquatic Biology Research Group, Hatherly Laboratories, University of Exeter, Prince of Wales Road, Exeter, EX4 4PS

^{‡‡‡}Division of Biology, Imperial College London, Silwood Park, Ascot SL5 7PY, United Kingdom

^{§§§}Great Lakes Water Institute, 600 East Greenfield Avenue, Milwaukee, Wisconsin 53204, USA

^{|||||}Environment Canada, 867 Lakeshore Road, Burlington, Ontario, Canada L7R 4A6

^{###}Ecotox Solutions, Schwarzwaldallee 215, CH-4058 Basel, Switzerland

^{††††}Fraunhofer IME, Auf dem Aberg 1, 57392 Schmallenberg, Germany

^{‡‡‡‡}CERI, 19-14, Chuo-machi, Kurume-shi, Fukuoka, 830-0023, Japan

^{§§§§}Federal Environment Agency UBA, Wörlitzer Pl 1, 06844 Dessau-Rosslau, Germany

^{||||||}Ecostat, Vondellaan 23, 2332 AA Leiden, The Netherlands

^{####}Deltares, PO Box 177, 2600 MH Delft, The Netherlands and Institute for Environmental Studies, VU University, De Boelelaan 1087, 1081 HV Amsterdam, The Netherlands

^{†††††}Centre for Ecology and Hydrology, Lancaster Environment Centre, Library Avenue, Bailrigg, Lancaster, Lancashire LA1 4AP, United Kingdom

^{‡‡‡‡‡}BASF SE, GV/TC - Z570, 67056 Ludwigshafen, Germany

^{§§§§§}Syngenta, Environmental Safety, Jealott's Hill International Research Centre, Bracknell, Berkshire, RG42 6EY, United Kingdom

(Submitted 7 October 2009; Returned for Revision 1 December 2009; Accepted 21 December 2009)

ABSTRACT

Fish full life cycle (FFLC) tests are increasingly required in the ecotoxicological assessment of endocrine active substances. However, FFLC tests have not been internationally standardized or validated, and it is currently unclear how such tests should best be designed to provide statistically sound and ecologically relevant results. This study describes how the technique of multi-criteria decision analysis (MCDA) was used to elicit the views of fish ecologists, aquatic ecotoxicologists and statisticians on optimal experimental designs for assessing the effects of endocrine active chemicals on fish. In MCDA qualitative criteria (that can be valued, but not quantified) and quantitative criteria can be used in a structured decision-making process. The aim of the present application of MCDA is to present a logical means of collating both data and expert opinions on the best way to focus FFLC tests on endocrine active substances. The analyses are presented to demonstrate how MCDA can be used in this context. Each of 3 workgroups focused on 1 of 3 species: fathead minnow (*Pimephales promelas*), Japanese medaka (*Oryzias latipes*), and zebrafish (*Danio rerio*). Test endpoints (e.g., fecundity, growth, gonadal histopathology) were scored for each species for various desirable features such as statistical power and ecological relevance, with the importance of these features determined by assigning weights to them, using a swing weighting procedure. The endpoint F1 fertilization success consistently emerged as a preferred option for all species. In addition, some endpoints scored highly in particular species, such as development of secondary sexual characteristics (fathead minnow) and sex ratio (zebrafish). Other endpoints such as hatching success ranked relatively highly and should be considered as useful endpoints to measure in tests with any of the fish species. MCDA also indicated relatively less preferred endpoints in fish life cycle tests. For example, intensive histopathology consistently ranked low, as did measurement of diagnostic biomarkers, such as vitellogenin, most likely due to the high costs of these methods or

* To whom correspondence may be addressed:

mark.crane@wca-environment.com

Published online 20 January 2010 in Wiley InterScience

(www.interscience.wiley.com).

DOI: 10.1002/ieam.43

their limited ecological relevance. Life cycle tests typically do not focus on identifying toxic modes and/or mechanisms of action, but rather, single chemical concentration–response relationships for endpoints (e.g., survival, growth, reproduction) that can be translated into evaluation of risk. It is, therefore, likely to be an inefficient use of limited resources to measure these mechanism-specific endpoints in life cycle tests, unless the value of such endpoints for answering particular questions justifies their integration in specific case studies. *Integr Environ Assess Manag* 2010;6:xxx–xxx. © 2010 SETAC

Keywords: Multi-criteria decision analysis MCDA Fish life cycle test Endocrine active substance

INTRODUCTION

Fish full life cycle (FFLC) tests are being included in conceptual frameworks for the assessment of endocrine active substances (e.g., OECD 2002; USEPA 2002). However, it is unclear how such tests should best be optimized in terms of design and degree of replication to provide ecologically relevant and statistically sound results, while minimizing cost and animal use. This study presents the outcome of a workshop organized for the CEFIC (European Chemical Industry Council) Long-Range Research Initiative (<http://www.cefic-lri.org/>). The workshop comprised an invited group of aquatic ecotoxicologists, fish ecologists and statisticians (the authors of this study).

This study uses data and structured expert judgment to identify which fish life cycle test endpoints are the most preferred for assessing the ecological risk of chemicals in general and endocrine active chemicals in particular. The question of endpoint selection is one that arises regularly in the field of ecotoxicology and is most logically addressed by considering endpoint sensitivity, statistical power, ecological significance, and cost. A common problem in answering it is that many ecotoxicologists pursue sensitivity of response without adequately considering power and ecological significance. On the other hand, those ecologists who take an interest in ecotoxicology may not always understand the expense, difficulty, and sometimes the technical impossibility, of running a fish study to detect small changes in highly variable endpoints. In essence, appropriate endpoints in fish life cycle tests must be 1) of biological importance at or associated with the population level, defined herein as ecologically relevant (and hence relevant for regulation), 2) sensitive to chemicals (and the level of effect which is biologically significant should be understood), and 3) of known statistical power for detecting biologically important levels of response. Omission of any of the above when deciding on appropriate endpoints is likely to lead to suboptimal choices. Finally, all these criteria need to be considered under overarching considerations of cost and animal ethics. The best test that could possibly be conducted is not always reasonable from a resource perspective.

The use of FFLC experiments specifically to evaluate the potential environmental impact of endocrine active chemicals is quite recent. Many effects of endocrine active chemicals are sub-lethal, for example altered sexual differentiation (Anderesen et al. 2003) and reduced fecundity (Nash et al. 2004). The complete FFLC test ensures that these effects are identified, allowing longer-term impacts on populations to be simulated. Most proposed test designs involve the addition of endocrine-relevant “mechanistic” endpoints to standard FFLC test designs, such as sex ratio, gonadal histopathology and vitellogenin induction (OECD 2008). It is probable that concern for endocrine-mediated effects in fish (most likely derived from findings in the mammalian toxicology database

or a positive result from an *in vivo* fish endocrine screening test) would be used to trigger a FFLC test with endocrine endpoints. This is likely to be the case under the revisions to the European Plant Protection Products Directive (91/414) and Registration, Evaluation, Authorisation and restriction of Chemicals (REACH; as substances of “equivalent concern”). Similarly, in the US the Endocrine Disrupter Screening Program (<http://www.epa.gov/endo/>) would require a FFLC study as a Tier II test following a weight of evidence evaluation of any positive indicators of endocrine activity at Tier I (screening). Consequently, the demand to conduct FFLCs to support risk assessments for endocrine-active chemicals may increase in the future.

Currently, FFLC tests have not been internationally standardized and validated (intercalibrated), which is not surprising, considering the length, cost, and complexity of the studies, although the US Environmental Protection Agency (USEPA) and other jurisdictions have published guidance documents (Hansen et al. 1978; Benoit 1981; Anon 2002) based largely on tests conducted before the need for risk assessments for endocrine-active substances. Validation is particularly important for FFLC tests because they are difficult, time consuming, animal intensive, and expensive to conduct. In addition to this, some endpoints (e.g., fecundity) measured in these tests, while frequently sensitive, are inherently variable, producing noisy data with relatively low statistical power (Crane and Matthiessen 2007). Any simplifications or improvements in fitness-for-purpose of test design and aids to interpretation of data generated would be highly desirable.

FFLC tests generally begin with fertilized eggs (F0 generation), which are continuously exposed to the test substance until the eggs have developed into adults, a proportion of which (while still exposed) are allowed to breed and produce offspring (F1), which are followed until at least the swim-up (early fry stage) or an early life stage (e.g., 30 d after hatch). However, different designs are now also under consideration. In some cases, life cycle tests are continued until the point where the F1 generation is sexually differentiated (e.g., an extended 1-generation test). Two-generation tests have also been conducted, as have a few multigeneration tests which continue until the F2 generation reaches swim-up (reviewed in OECD 2008). These extended designs have been proposed to address the potential issues of maternal transfer of strongly bioaccumulative substances or endocrine-mediated transgenerational effects. There is limited evidence from a recent Detailed Review Paper (DRP) by the Organization for Economic Cooperation and Development (OECD) that prolonged tests of this type might be more sensitive to strongly bioaccumulative substances (OECD 2008). The evidence for transgenerational effects in these fish tests is also limited.

Although the FFLC test might be considered definitive in ecotoxicological terms (by covering all portions of the life

cycle), it should be recognized that the test does not precisely mimic reproduction, as would occur in the field (particularly for species with reproductive strategies which differ strongly from those used in tests), and, therefore, the results may be difficult to apply directly to field conditions. For example, species used for most FFLC tests perform best as pairs or small breeding groups, but many fish species (e.g., roach, *Rutilus rutilus*) breed in much larger groups, and are not constrained by the limited space available under laboratory conditions. There is also evidence that sperm release in some species (salmonids) is very sensitive to chemical interference, with the male olfactory epithelium able to detect female pheromonal prostaglandins (e.g., Moore and Lower 2001), but it is doubtful whether current FFLC tests take this process into account fully, or at all. Breeding behavior in the laboratory may, therefore, not be representative of some fish species in the wild, which could result in an important uncertainty given that some endocrine active chemicals alter sexual behavior (e.g., Martinović et al., 2007; Maunder et al. 2007). Laboratory fish also tend to be of limited genetic background when compared with their wild conspecifics (Coe et al. 2009), which could mask the individual variations in sensitivity that can occur. Fish in confined test chambers will experience an altered degree of stress in captivity compared with the field, as well as artificial social hierarchies, potentially modifying an individual's sensitivity to toxicants. Conversely, wild fish populations may be subject to multiple "background" variables or stressors (aside from toxicity due to any particular chemical), which contribute to environmental stochasticity and the possibility of population decline (Brown et al. 2003). Environmental stressors (e.g., temperature fluctuations, competition, and predation) are absent in most laboratory life cycle fish exposures. Competition (for mates, food, or space) is not taken into account when fish are paired or placed in mating groups, and factors such as predation (of adults or young) are missing entirely. Finally, domestication effects lead to divergence of many life history traits between wild and cultured or laboratory populations (Thorpe 2004; Lorenzen 2005). However, these laboratory-to-field extrapolation uncertainties are accounted for, at least to some extent, in the risk assessment evaluation procedure by the use of assessment (safety) factors applied to the data derived from FFLC tests.

Published FFLC test data on endocrine active chemicals are sparse, and predominantly involve just 3 species: the Japanese medaka (*Oryzias latipes*), the fathead minnow (*Pimephales promelas*), and the zebrafish (*Danio rerio*). OECD (2008) lists 15, 3, and 7 life cycle tests with endocrine active chemicals for each species, respectively. In that document there are a further 16 life cycle test datasets for nonendocrine active chemicals with, predominantly, the fathead minnow, and limited additional data for sheepshead minnow (*Cyprinodon variegatus*), flagfish (*Jordanella floridae*), brook trout (*Salvelinus fontinalis*), medaka, and zebrafish.

An additional uncertainty is the fact that few field-based experiments with fish populations exposed to endocrine active chemicals have been conducted, so the ability of FFLC tests to predict effects at the population level is largely unknown. Furthermore, the question whether natural fish populations have actually been damaged by endocrine active chemicals remains unresolved (Mills and Chichester 2005), despite abundant evidence that wild fish have *inter alia* been feminized by exposure to estrogens (e.g., World Health

Organization 2002; Matthiessen 2006). A key reason for this apparent discrepancy may be the capacity of fish populations to compensate for reductions in reproductive output, survival or growth through density-dependent increases in other vital rates (Rose et al. 2001). The strength of compensatory processes differs between populations, but in general compensatory reserve is greater in juvenile survival than in adult growth or reproductive traits (Lorenzen 2008). This suggests that population abundance is more sensitive to ecotoxicological effects on adult life history traits than to effects on juvenile traits, with the exception of effects on sex ratio, which are caused during juvenile exposure. Consequently, test endpoints relating to adult growth, survival and reproductive traits (including those that are caused during juvenile sexual development) are particularly relevant to population level effects. In large populations in which density dependence acts mostly in a compensatory manner, effects on population abundance will generally be smaller than effects measured in the laboratory for particular life stages and vital rates. However, small populations may show density dependence (Allee effects; Allee 1931), such that ecotoxicological reductions in vital rates may be larger and possibly lead to catastrophic effects on population abundance.

One example for which limited field data exist concerns the synthetic estrogen ethynylestradiol (EE2), which is widespread in sewage effluents and surface waters at low ng/L concentrations. Kidd et al. (2007) conducted a whole-lake dosing experiment in Canada, in which EE2 was added to Lake 260 every year from May to October for 3 y, producing mean annual measured concentrations in epilimnetic waters of 6.1, 5.0, and 4.8 ng/L over the 3 years of dosing. Recruitment of fathead minnows substantially failed in the third year, possibly due to reduced fertility, and the population had almost disappeared by year 6 (i.e., 3 y after dosing ceased). It is instructive to compare these Canadian field data with laboratory-based life cycle tests that have been conducted with fathead minnows exposed to EE2 (Länge et al. 2001; Parrott and Blunt 2005), which have generated Lowest Observed Effect Concentrations (LOECs) for abnormal testicular development, fertilization success and sex ratio in the range <0.32-4 ng/L. Similar results have been obtained with the medaka (Balch et al. 2004), the zebrafish (Wenzel et al. 2001; Van den Belt et al. 2003), and the Chinese rare minnow *Gobiocypris rarus* (Zha et al. 2008). By constructing a life table based on the vital rates of survival and fecundity for fathead minnows from the study by Länge et al. (2001), Grist et al. (2003) used a Leslie matrix model to show that EE2 concentrations of 0.53-3 ng/L would reduce the intrinsic rate of population increase (r) by 20% and 100% compared with the control. In broad terms, therefore, this demonstrates reasonably good correspondence between the results of FFLC and field experiments, as confirmed by agreement with a Predicted No Effect Concentration (PNEC) of 0.35 ng/L derived from a species sensitivity distribution approach (Caldwell et al. 2008).

It is probable that FFLC tests will become part of a suite of tools to assess the potential ecological risk of endocrine-active chemicals. However, there are many options in terms of experimental design, species and endpoints. Unfortunately, little comparative information exists on which to base recommendations for appropriate FFLC test designs. This study describes how the technique of multi-criteria decision analysis (MCDA) was used to elicit the views of fish

ecologists, aquatic ecotoxicologists, and statisticians on optimal experimental designs for assessing the effects of endocrine active chemicals on fish. In MCDA qualitative criteria (that can be valued, but not quantified) and quantitative criteria can be used in a structured decision-making process (DTLR 1999). Kiker et al. (2005) describe how MCDA can be used for environmental decision making, and Yatsalo et al. (2007) provide a practical example of this. The aim of the present application of MCDA is to present a logical means of collating both data and expert opinions on the best way to focus FFLC tests on endocrine active substances. The following analyses are presented to demonstrate how MCDA can be used in this context. We should point out that MCDA is only 1 of a range of potential structured approaches to environmental decision making. The results should be regarded as preliminary and designed to illustrate the process. They can certainly be refined further with additional input.

METHODS

A 4-d workshop was convened in Palma, Mallorca, Spain, 23–26 September 2008, to which 25 fish ecologists and ecotoxicologists were invited, along with experts in statistics and environmental regulation. The first part of the workshop was spent considering what evidence is available on the relationship between fish reproduction and population-level effects in laboratory and field studies. The results of sensitivity and power analyses from Crane and Matthiessen (2007) were also presented at the workshop.

The workshop participants were asked to consider the evidence and then provide information for use in an MCDA to determine what endpoints in fish tests provide the optimum balance between sensitivity, power and ecological significance, at a study size and cost that is practical to implement. First they were asked to construct a “value tree,” which is a graphical representation of the different criteria that they wanted to use to appraise their different decision options. The value trees could be used to organize criteria into tiers (called subnodes) if required for clarity of analysis. The groups were then asked to score each of the options against each of the criteria, either directly on a preference scale of 0 to 100 or on an equivalent natural scale (e.g., cost) that was subsequently converted to a scale of 0 to 100. Finally they used the technique of swing weighting to assign weights to each of the criteria (Belton and Stewart 2002). This approach determines how a swing in weight from 0 to 100 for 1 criterion compares with a similar swing for another criterion. For example, swing weighting can be used to decide whether a difference in the range of cost per study endpoint from, say, \$20 000 (least expensive, so scored 100) to \$24 000 (most expensive, so scored 0) is more or less important than a difference in number of fish used per study endpoint of 1000 (lowest number, so scored 100) to 2000 (highest number, so scored 0). Individuals asked to consider this question might believe that cost per se is an important criterion when choosing between endpoints, but that the difference in cost between the least expensive and most expensive endpoints in this example is rather insignificant when compared with the difference between numbers of animals used to determine these endpoints.

The participants performed these tasks in 3 separate breakout groups—1 for each of the main small fish species used in FFLC tests (zebrafish, medaka, and fathead minnow,

with each group comprising a selection of fish ecologists, ecotoxicologists, statisticians, and environmental regulators). These groups then reported back to plenary sessions. After considering the views of the other groups, the experts were asked to revisit their conclusions and produce a final view. This 2-stage approach is valuable, because it is common for there to be a learning process for participants during MCDA (Sparling and Tarbotton 2000).

The MCDA software program HiView Version 3.2 (Catalyze, Winchester, UK) was used to collate and analyze the data and produce the following outputs for each group: 1) A value tree, 2) A table of the final scores and weights, which identified the preferred endpoints and the criteria that contributed to these preferences, 3) Sensitivity analyses to determine what percentage change in weight would be required to change the preferred endpoint to an alternative endpoint. These changes were banded <5%, 5–15%, and >15% change in weight.

Fish test endpoints considered for this evaluation included the following: Time to hatch; Hatching success (number of embryos that complete hatching, expressed as a percentage of eggs deemed fertile); Fry survival; Growth; Condition factor ($[\text{weight} \times 100] / \text{length}^3$); Sex ratio (macroscopic observation and histological confirmation); Sexually undifferentiated ratio (fish that have not sexually matured); Secondary sexual characteristics (e.g., fatpad and tubercles in fathead minnows, and anal fin shape and papillary processes in medaka, which are under endocrine control and can be enhanced, suppressed or induced); gonadosomatic index (gonad weight/body weight $\times 100$); Gonad sex determination (sex determination based on gonads, rather than external appearance); Major histological abnormalities (e.g., intersex is the development of both ovarian and testicular tissues in the gonads); Intensive histology (a variety of alterations in gonad histology that may be associated with exposure to an endocrine active substance, and intensive techniques such as staging and potentially assessment of multiple tissues); Vitellogenin induction and/or reduction (a female egg yolk protein that can be induced in males following exposure to estrogenic substances or reduced in females in response to estrogen antagonists); Time to spawn (age at spawning or surrogate, i.e., time elapsed during study before spawning); Behavior (reproductive behavior, e.g., loss of territorial aggressiveness or spawning behavior); Pheromone production; Fecundity (number of viable eggs produced by females); Fertilization success (the number of fertile eggs, expressed as a percentage of the number of total eggs); and Population growth (expressed as the intrinsic rate of increase).

RESULTS

Fathead minnow

The fathead minnow group developed the final value tree shown in Figure 1. They identified 24 endpoints that could logically be measured in fathead minnow FFLC tests with endocrine-active chemicals: Time to hatch (F0, F1); Hatching success (F0, F1); Fry survival (F0, F1); Growth (F0, F1); Secondary sexual characteristics (F0, F1); Gonadosomatic index (F0, F1); Gonad sex determination (F0, F1); Major histological abnormalities (F0, F1); Intensive histology (F0, F1); Vitellogenin induction (F0, F1); Time to spawn (F0); Behavior (F0); Fecundity (F0); and Fertilization success (F1).

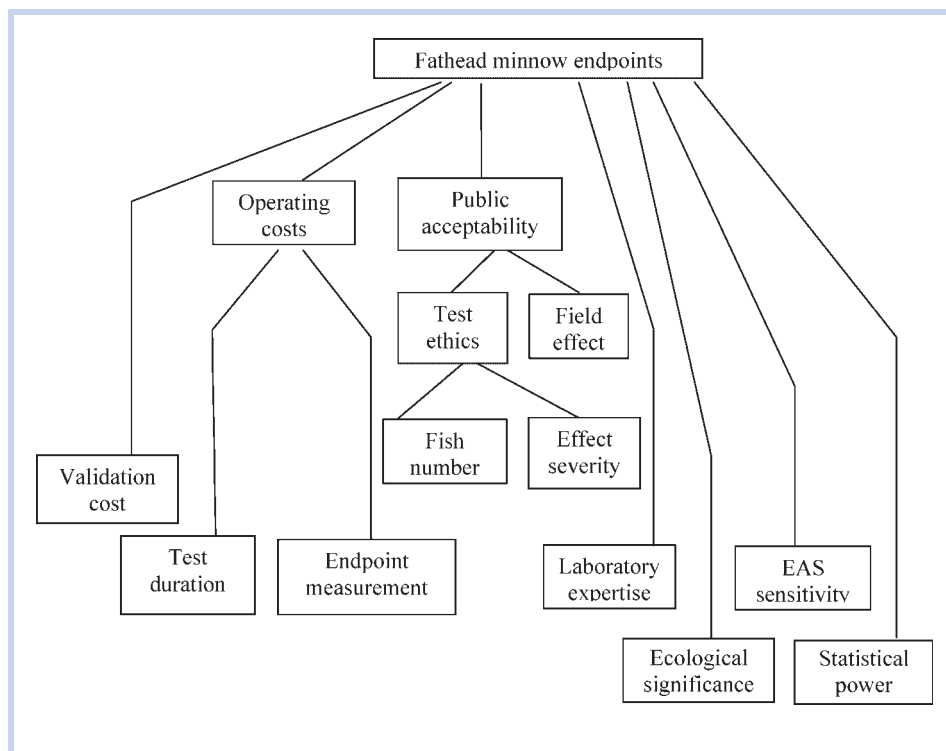


Figure 1. Value tree for criteria when selecting fathead minnow endpoints. EAS = endocrine active substance.

Scores for endpoints against the criteria in Figure 1, plus criteria weights derived by the group through the technique of swing weighting are shown in Table 1. The values in Table 1 (and also in Tables 2 and 3) should be interpreted as follows. The different potential FFLC test endpoints identified by the group are listed in the top row, and the criteria used to choose between these endpoints are listed in the first column. Values in bold are for higher node criteria (Tier 1) in the value tree that integrate any criteria below. Values in normal font are for Tier 2 criteria that are integrated by tier 1, and values in italics are for tier 3 criteria that are integrated by tier 2 (only the fathead minnow group decided to include tier 3 criteria). The scores from 0 to 100 given by the group to each endpoint against each criterion are shown in the columns under each endpoint. Finally, the swing weights given to each of the criteria are shown in column 2. Note that swing weights given to lower nodes propagate to higher nodes where they could again be checked for plausibility by the group.

The results in Table 1 show that the group weighted relative ecological significance of the different endpoints most highly, with differences in endpoint sensitivity to endocrine active substances weighted almost as highly. All other criteria (e.g., operating costs, statistical power, required level of laboratory expertise) for selecting endpoints were weighted as being considerably less important. F1 fertilization success emerged as the endpoint with the highest overall score, when criteria scores and weights were combined. This is because this endpoint scored relatively highly across several criteria, including those that were weighted most highly by the group. Other endpoints that scored highly overall were F0 and F1 secondary sexual characteristics, and F0 fecundity.

Sensitivity analysis showed that the most preferred endpoint of F1 fertilization success is sensitive to relatively

small changes in weighting of less than 5%. Less than a 5% increase in weighting on test duration or field effect, or a 5–15% increase in weighting on endpoint measurement and number of fish used, would change the most preferred endpoint to F0 secondary sexual characteristics. A 5–15% decrease in weighting on laboratory expertise would change the most preferred endpoint to F0 secondary sexual characteristics, while a similar decrease in weighting on endocrine active chemical sensitivity and statistical power would change the most preferred option to F0 time to hatch and F0 fecundity, respectively.

The lowest scoring endpoints were F0 and F1 intensive histopathology because of relatively low scores against all of the criteria.

In summary, for fathead minnow, the workshop group identified F1 fertilization success, F0 fecundity, F0 time to hatch, and F0/F1 secondary sexual characteristics as the most preferred endpoints in fathead minnow FFLC tests, primarily on the basis of their high ecological significance (see *Discussion* section for further assessment) and their sensitivity to endocrine active substances.

Medaka

The medaka group developed the final value tree shown in Figure 2. They identified 25 endpoints measured in medaka FFLC tests: Time to hatch (F0, F1); Hatching success (F0, F1); Fry survival (F0, F1); Growth (F0, F1); Sex ratio (F0, F1); Major histological abnormalities (F0, F1); Vitellogenin induction (F0, F1); Behavior (F0, F1); Population growth (F0, F1); Pheromone production (F0); Time to spawn (F0); Fecundity (F0); Condition factor (F0); Gonadosomatic index (F0, F1); and Fertilization success (F1).

Table 1. MCDA scores and weights for fathead minnow FFLC endpoints. Total scores and weights are for the highest nodes (tier 1) in the value tree. Subnodes that contribute to the main nodes are in normal text (tier 2) or italicized text (tier 3).

Criteria	Criteria weights	F0 time to hatch	F0 hatching success	F0 fry survival	F0 growth	F0 ² sexual characteristics	F0 gonadosomatic index	F0 gonad sex determination	F0 major histological abnormalities	F0 intensive histology	F0 vitellogenin induction	F0 time to spawn	F0 behavior	F0 fecundity	F1 fertilization success	F1 time to hatch	F1 hatching success	F1 fry survival	F1 growth	F1 ² sexual characteristics	F1 gonadosomatic index	F1 gonad sex determination	F1 major histological abnormalities	F1 intensive histology	F1 vitellogenin
Validation costs	10	100	100	100	100	80	100	80	60	40	80	80	0	75	100	100	100	100	100	80	100	80	60	40	80
Operating costs	20	100	100	88	52	52	52	40	34	28	43	50	32	18	18	30	30	30	24	24	24	12	6	0	15
Test duration	14	100	99	83	40	40	40	40	40	40	40	29	29	0	0	0	0	0	0	0	0	0	0	0	0
Endpoint measurement	6	100	100	100	80	80	80	40	20	0	50	100	40	60	60	100	100	100	80	80	80	40	20	0	50
Public acceptability	40	79	84	14	55	65	55	65	65	55	31	55	72	60	51	46	51	5	46	56	46	56	56	46	22
Field effects	4	0	50	50	0	100	0	100	100	0	100	0	50	50	50	0	50	50	0	100	0	100	100	0	100
Test ethics	36	87	87	10	61	61	61	61	61	61	23	61	74	61	51	51	51	0	51	51	51	51	51	51	13
Fish number	13	100	29	29	29	29	29	29	29	29	29	29	29	29	0	0	0	0	0	0	0	0	0	0	0
Severity	23	80	80	0	80	80	80	80	80	80	20	80	100	80	80	80	80	0	80	80	80	80	80	80	20
Laboratory expertise	40	100	100	100	100	80	100	60	40	20	70	100	0	100	100	100	100	100	100	80	100	60	40	20	70
Ecological significance	100	80	100	100	80	90	70	80	30	0	0	80	90	100	100	80	100	100	80	90	70	80	30	0	0
EAS sensitivity	90	0	0	0	20	90	40	90	100	60	100	20	80	100	90	0	0	0	20	90	40	90	100	60	100
Statistical power	50	100	0	0	0	50	50	0	50	50	50	100	50	0	50	100	50	0	0	50	50	0	50	50	50
TOTAL	350	66	65	50	52	78	61	65	57	34	49	65	63	76	80	58	58	45	49	75	58	62	54	31	46

Table 2. MCDA scores and weights for medaka FFLC endpoints. Total scores and weights are for the highest nodes (tier 1) in the value tree. Subnodes that contribute to the main nodes are in normal text (tier 2) or italicized text (tier 3).

Criteria	Criteria weights	F0 time to hatch	F0 hatching success	F0 fry survival	F0 growth	F0 sex ratio	F0 major histological abnormalities	F0 time to spawn	F0 fecundity	F0 condition factor	F0 gonadosomatic index	F1 fertilization success	F1 time to hatch	F1 hatching success	F1 fry survival	F1 growth	F1 sex ratio	F1 major histological abnormalities	F0/F1 Biomarker effects (vitellogenin)	F0/F1 Behavioral effects	F0/F1 Pheromone effects	F0/F1 Population level effects (e.g.,r)
Financial	20	100	100	90	85	65	50	66	64	78	79	39	39	39	21	16	15	0	78	78	74	37
Fish number	4	100	100	100	100	100	100	100	100	100	100	0	0	0	0	0	0	0	100	100	100	100
Technician time	6	100	100	88	88	24	24	24	44	68	68	36	36	36	8	0	0	0	68	68	68	8
Lab costs	10	100	100	86	77	74	46	78	62	76	77	57	57	57	38	31	30	0	76	76	68	30
Non financial	260	61	66	61	58	59	55	57	50	51	36	78	61	65	59	46	58	51	29	28	8	61
Ecological relevance	100	60	93	87	73	73	47	93	100	73	7	100	60	93	87	73	60	33	0	53	13	100
Public explainability	10	21	37	37	79	100	89	37	68	58	95	68	5	37	26	37	96	89	0	26	0	23
Statistical power	90	100	74	59	49	49	59	29	0	39	39	74	100	74	59	23	49	59	29	18	8	13
EAS sensitivity	50	0	0	18	35	41	53	35	41	35	82	54	13	5	18	35	71	71	100	0	0	71
Risk of false positives	10	67	67	67	83	67	67	67	17	17	17	17	17	17	17	33	17	17	0	0	0	100
TOTAL	280	64	68	63	60	60	54	57	51	53	39	75	59	63	56	44	55	47	33	31	13	60

Table 3. MCDA scores and weights for zebrafish FFLC endpoints. Total scores and weights are for the main nodes in the value tree (bold text). Total scores and weights are for the highest nodes (tier 1) in the value tree. Subnodes that contribute to the main nodes are in normal text (tier 2).

Criteria	Criteria weights	Biomarkers	Histopathology	Time to spawn	Fecundity	Fertilization success	Time to hatch	Hatching success	Sex ratio	Undifferentiated ratio	Survival	Length	Weight	Behavior	Overall "fitness"
Costs	220	41	27	49	66	87	65	82	74	68	82	46	53	67	64
Study duration	10	100	0	100	100	100	100	100	25	25	100	100	100	100	100
Operating costs	40	71	0	94	63	46	89	89	89	89	85	82	100	96	89
Laboratory capability/capacity	10	38	6	75	88	75	88	88	25	25	100	75	100	25	0
False positive results	50	85	69	92	100	100	92	100	92	8	100	0	23	100	100
False negative results	50	0	46	15	77	92	85	92	31	92	92	100	92	92	92
Public's views	60	10	0	0	20	100	0	50	100	100	50	0	0	0	0
Benefits	310	45	41	71	60	75	39	50	94	69	56	82	82	38	28
Specificity to EAS	20	100	100	10	10	50	0	0	100	100	0	10	10	80	0
Sensitivity to EAS	70	0	20	60	60	60	60	73	100	20	100	73	73	0	0
Ethical considerations	5	100	100	100	100	100	100	100	100	100	100	100	100	100	0
Ability to extrapolate across environments	95	100	80	100	40	80	0	0	100	80	0	100	100	50	13
Regulatory relevance	20	100	0	0	0	0	0	0	0	0	0	0	0	0	0
Population relevance	100	0	13	75	100	100	75	100	100	100	100	100	100	50	75
TOTAL	530	43	35	62	63	80	50	64	85	69	67	67	70	50	43

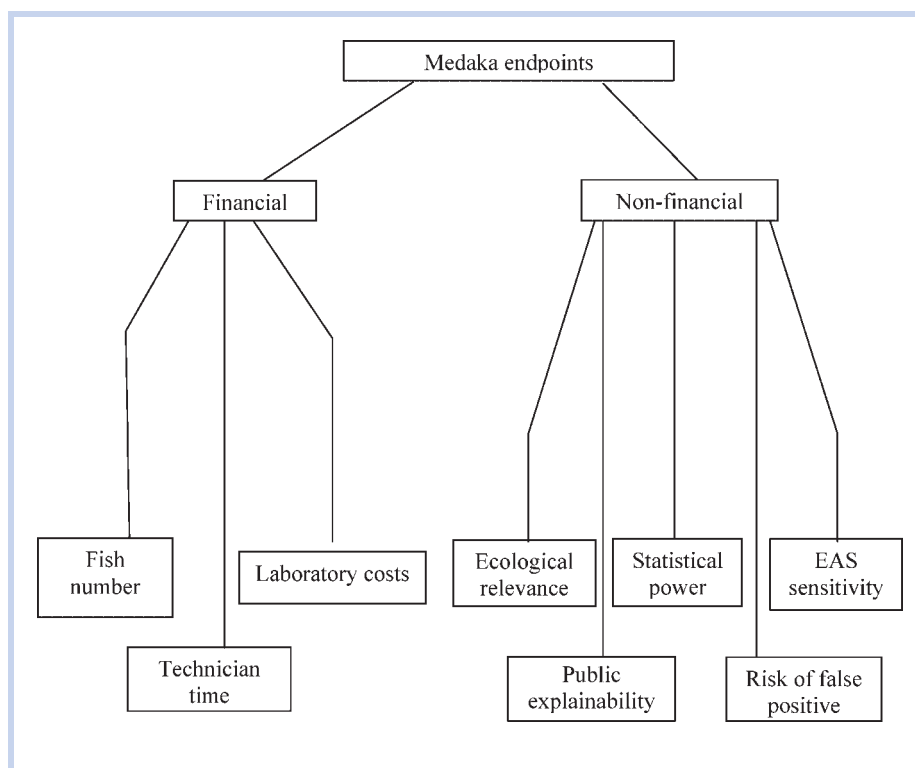


Figure 2. Value tree for criteria when selecting medaka endpoints. EAS = endocrine active substance.

Scores for endpoints against the criteria in Figure 2, plus criteria weights applied by the group are shown in Table 2. This shows that, like the fathead minnow group, the medaka group weighted ecological significance between the different endpoints most highly, with differences in statistical power being weighted almost as highly. All other criteria for selecting endpoints were weighted as being considerably less important.

F1 fertilization success emerged as the endpoint with the highest overall score, when criteria scores and weights were combined, which was also the case for the fathead minnow group. Other endpoints that scored relatively highly overall were F0 and F1 hatching success. Sensitivity analysis showed that the most preferred endpoint of F1 fertilization success is sensitive to changes in weighting of 5–15%. A 5–15% increase in weighting on fish numbers, technician time or laboratory costs would change the most preferred endpoint to F0 hatching success. A 5–15% decrease in weighting on endocrine active substance sensitivity would also change the most preferred endpoint to F0 hatching success. This is because the latter endpoint is generated at the beginning of the study and is attractive because of its low operational cost.

The lowest scoring endpoints were effects on vitellogenin induction, behavior and pheromones, although in some cases (e.g., pheromone production) it was argued by group members that this could be due to a lack of experience with and knowledge of these endpoints.

In summary, for medaka, the workshop group identified F1 fertilization success and F0 and F1 hatching success as the most preferred endpoints in medaka FFLC tests, primarily on the basis of their relatively high ecological significance and statistical power (see *Discussion* section for further assessment). The higher ranking of F1 fertilization success was also

based on the understanding that the cost of a FFLC study is outweighed by the importance of completing the study and product approval process on schedule and avoiding any delay to market.

Zebrafish

The zebrafish group developed the final value tree shown in Figure 3. In contrast to the other 2 groups, this group combined the evaluation of endpoints in F0 and F1 generations, to prevent F1 endpoints being unfairly affected by certain criteria. This resulted in 14 combined generic endpoints: Biomarkers (vitellogenin and the male hormone, 11-ketotestosterone, plus specific biomarkers for other endocrine modes of action if these emerge in the future); Histopathology (full body, i.e., multiple organs); Time to spawn; Fecundity; Fertilization success; Time to hatch; Hatching success; Sex ratio; Sexually undifferentiated ratio; Survival; Length; Weight; Behavior (e.g., spawning behavior); and “Fitness.” (The group tried to consider endpoints which may be of ecological relevance but are currently not considered or tested yet. An example is the influence of a chemical on the response to stress, which could be tested as a laboratory challenge test, in which the organisms react to stimuli. This type of endpoint was referred to as “fitness.”)

Scores for endpoints against the criteria in Figure 3, plus criteria weights applied by the group are shown in Table 3. This shows that, like the other groups, the zebrafish group weighted the differences in ecological significance (which they called “population relevance”) between the different endpoints most highly, with differences in the ability to extrapolate results across different environments being

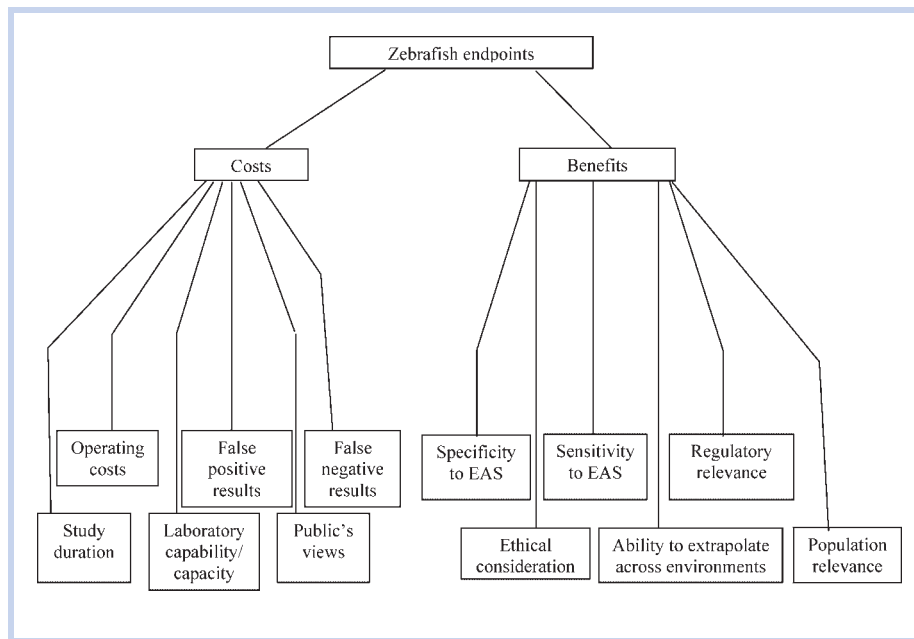


Figure 3. Value tree for criteria when selecting zebrafish endpoints. EAS = endocrine active substance.

weighted almost as highly. Sensitivity to endocrine active substances also received a relatively high rating. All other criteria for selecting endpoints were weighted as being considerably less important.

Sex ratio emerged as the endpoint with the highest overall score, when criteria scores and weights were combined. This was mainly due to the recognized and irreversible effects of endocrine active chemicals on reproductive organs (e.g., Nash et al. 2004) and the potential negative effects that would result at the population level. However, it was noted that the genetic basis of sex determination in zebrafish is only partially understood (Jørgensen et al. 2008) and sex ratio may be variable even under controlled laboratory conditions according to recent ring test data for the Fish Sexual Development Test. Plasticity in sex ratio appears to depend on numerous environmental factors, including food availability and temperature during critical early development and also degree of inbreeding versus outbreeding (see Lawrence et al. 2008).

In common with medaka and fathead minnow, the other endpoint that scored relatively highly overall was F1 fertilization success. Sensitivity analysis showed that the most preferred endpoint of sex ratio is sensitive to changes in weighting of 5–15%. A 5–15% increase in weighting on study duration, laboratory capability or capacity, or false negative results would change the most preferred endpoint to F1 fertilization success. A 5–15% decrease in weighting on regulatory relevance would also change the most preferred endpoint to F1 fertilization success. The lowest scoring endpoints were effects on biomarkers, histopathology and overall fitness.

In summary, for zebrafish, the expert group identified sex ratio and F1 fertilization success as the most preferred endpoints in zebrafish FFLC tests, primarily on the basis of their relatively high ecological significance (see *Discussion* section for further assessment), the ability to extrapolate results across different environments, and sensitivity to endocrine active substances.

DISCUSSION

It is interesting to note that the 3 groups constructed different value trees. This does not appear to be due to any basic differences in the FFLC tests under consideration, but instead reflects differences in the views of individuals within each group about how to frame the overall question. This illustrates the importance of taking the existence of such different views transparently into account, even when a decision is being informed by apparently objective experts.

However, despite developing initial value trees, criteria, scores and weightings independently, all 3 workshop groups arrived at rather similar overall conclusions. Fertilization success (F1) emerged as a high priority measurement endpoint from all groups. However, it became apparent subsequent to the workshop that participants were using different definitions of the term fertilization success. As a result, this endpoint scored highly because it was considered to be both ecologically relevant and of high statistical power, when in fact each of these is an attribute of 2 different definitions. Fertilization success, defined as the fertile proportion of the total number of eggs, does have high statistical power, because it is expressed as a percentage and is, therefore, normalized. However, this definition of fertilization success is not associated with high ecological relevance (e.g., a fertilization success of 90% could be 9 of 10 or 900 of 1000 eggs). The other definition used, which more accurately may be termed fertility, is the number of fertile eggs (e.g., the value of 9 or 900 eggs in the previous example, rather than the proportion). This endpoint does have high ecological relevance, but is of low statistical power, because it has been shown to be a highly variable endpoint. Fish full lifecycle studies should, therefore, be designed to gather data on both endpoints as efficiently and effectively as possible (as one is a function of the other). In terms of showcasing how MCDA can be used in this context, this provides a good example of how definitions need to be very clear at the outset so that all participants feed into the process with a common understanding.

Some other preferred endpoints identified by the MCDA scored highly only in particular species, such as development of secondary sexual characteristics in fathead minnow and sex ratio in zebrafish. Indeed, sex ratio in zebrafish was considered by that group to be the most preferred endpoint, mainly due to the recognized effects of endocrine active chemicals on reproductive organs (e.g., Nash et al. 2004). Although not yet possible in zebrafish, the ability to determine sex genetically and compare this with the phenotypic sex greatly improves the statistical power and certainty in an effect. It could also reduce the number of excess fish which would otherwise be required to ensure equal numbers of males and females. Genetic sex probes already exist for the medaka (Matsuda et al. 2002; Nanda et al. 2002) and there is also the d-rR strain that possesses sex-linked pigmentation, which distinguishes genotypic sex (Aida 1921). Recent research in the United States and Denmark suggests that measurement of genetic sex in fathead minnow may soon be possible (A. Olmstead and G. Ankley, USEPA, personal communication). Other relatively highly preferred endpoints such as hatching success are also likely to be worth measuring in tests with all fish species.

The different groups were also in broad agreement about the least preferred endpoints in FFLC tests. Intensive histopathology consistently ranked low, as did measurement of biomarkers such as vitellogenin. These types of diagnostic endpoints can lend important insights into toxic mechanism of action, but are typically of less utility for predicting adverse effects in individuals or populations. Research has been undertaken recently to address this shortcoming and these endpoints may have potential in the future long-term use of FFLCs (Miller et al. 2007). Currently, in tiered testing programs for endocrine-active chemicals, diagnostic biomarker-type endpoints initially are used to flag chemical mechanisms of concern, while high-tier (e.g., FFLC) tests are intended to generate the type of population-relevant (reproduction) data needed for determination of ecological risk. This type of emphasis reflects an approach in which biomarkers are used as supporting evidence rather than directly in the risk assessment procedure (Hutchinson et al. 2006). It is, therefore, likely to be an inefficient use of available resources to require routine measurement of these endpoints in all FFLC studies, whose purpose is to ascertain population relevant impacts. That being said, in specific case studies or for specific regulatory requirements, the integration of population relevant endpoints and diagnostic biomarkers may be justified.

SUMMARY AND CONCLUSIONS

In summary, using the logical framework of MCDA, a limited number of preferred FFLC test endpoints for assessing endocrine active substances were identified by 3 workshop groups with expertise in statistics, fish ecology and aquatic toxicity testing with fathead minnow, medaka, and zebrafish. Effects on F1 fertilization rate emerged as a generally preferred endpoint for all species, and also highlighted the need for clear definitions and a common understanding of all terms used in MCDA. Some endpoints scored highly for particular species, such as development of secondary sexual characteristics in fathead minnow and sex ratio in zebrafish. Indeed, sex ratio in zebrafish was considered by the group evaluating that species to be the most preferred endpoint due to its environmental (population) relevance. Other relatively

highly preferred endpoints such as hatching success would also be worth measuring in tests with any of the fish species.

There was also broad agreement about the least preferred endpoints in FFLC tests. Whereas histological confirmation of gonadal sex was considered important, intensive histopathology consistently ranked much lower, as did measurement of biomarkers such as vitellogenin. It is, therefore, likely to be an inefficient use of available resources to insist on routine measurement of these latter endpoints.

In conclusion, our analyses are preliminary and could certainly be refined further through discussion with a wider range of stakeholders, particularly to determine what swing weights should be allocated to criteria. For certain criteria there was considerable disagreement (e.g., public acceptability of testing) and further discussion matched with a public questionnaire would enable more uniform weighting to this criterion to be assigned. A decrease in the variability of swing weight scoring for this and other criteria would improve the MCDA results further. However, the examples presented here illustrate how MCDA can be used to provide a logical framework for scientists and regulators in which different options are identified, the criteria for choosing between them are discussed, scored, and weighted, and the sensitivity of the final results can be judged. In an economic environment in which government and industry scientists are continually required to do more with fewer resources, MCDA is a tool that can help identify essential and optimal ways forward. A clear and transparent audit trail is available when the MCDA process is followed, allowing anyone to return to and amend input data in the light of further scientific data, or different opinions about subjective criteria. This stands in contrast to the often opaque and untraceable outputs that can result from unstructured "expert judgment."

Acknowledgment—This work was funded by the Cefic Long Range Research Initiative (Project EMSG 47). The conclusions and recommendations made in this study reflect the views of the authors as individual scientists and do not represent a position of the organizations to which the authors are affiliated. We thank 2 anonymous peer reviewers for helpful comments on the original manuscript.

REFERENCES

- Aida T. 1921. On the inheritance of color in a fresh-water fish *Aplocheilichthys latipes* Temminck and Shlegel, with special reference to the sex-linked inheritance. *Genetics* 6:554–573.
- Allee WC. 1931. *Animal aggregations: A study in general sociology*. Chicago (IL): University of Chicago Press.
- Andersen L, Holbech H, Gessbo A, Norrgren L, Petersen GI. 2003. Effects of exposure to 17 alpha-ethinylestradiol during early development on sexual differentiation and induction of vitellogenin in zebrafish (*Danio rerio*). *Comp Biochem Physiol C* 134:365–374.
- Anon. 2002. The medaka (*Oryzias latipes*) full life cycle test guideline. Japan: Ministry of the Environment. 14 p.
- Balch GC, Mackenzie CA, Metcalfe CD. 2004. Alterations to gonadal development and reproductive success in Japanese medaka (*Oryzias latipes*) exposed to 17 alpha-ethinylestradiol. *Environ Toxicol Chem* 23:782–791.
- Belton V, Stewart TJ. 2002. *Multiple criteria decision analysis: An integrated approach*. Boston (MA): Kluwer Academic Publishers.
- Benoit DA. 1981. *User's guide for conducting life cycle chronic toxicity testing with fathead minnows (*Pimephales promelas*)*. US Environmental Protection Agency. EPA 600/8-81-011, NTIS / PB82-238395.

- Brown AR, Riddle AM, Cunningham NL, Kedwards TJ, Shillabeer N, Hutchinson TH. 2003. Predicting the effects of endocrine disruptors on fish populations. *Hum Ecol Risk Assess* 9:761–788.
- Caldwell DJ, Mastrocco F, Hutchinson TH, Länge R, Heijerick DG, Janssen CR, Anderson PD, Sumpter JP. 2008. Derivation of an aquatic predicted no-effect concentration for the synthetic hormone, 17 α -ethinyl estradiol. *Environ Sci Technol* 42:7046–7054.
- Coe TS, Hamilton PB, Griffiths AM, Hodgson D, Wahab MA, Tyler CR. 2009. Genetic variation in strains of zebrafish (*Danio rerio*) and the implications for ecotoxicology studies. *Ecotoxicology* 18:144–150.
- Crane M, Matthiessen P. 2007. An evaluation of the reproducibility and power of various endpoints in partial and full fish life cycle tests. Organisation for Economic Cooperation and Development, unpublished paper submitted to the 6th meeting of the OECD Validation Management Group for Ecotoxicity Tests. 2008 Jan 16–17. Paris. 90 p.
- DTLR. 1999. Multi-criteria analysis manual [Internet]. London (UK): Department of Transport, Local Government and Regions. Available from: www.communities.gov.uk/documents/corporate/pdf/146868.pdf
- Grist EP, Wells NC, Whitehouse P, Brighty G, Crane M. 2003. Estimating the effects of 17 α -ethinylestradiol on populations of the fathead minnow *Pimephales promelas*: Are conventional toxicological endpoints adequate? *Environ Sci Technol* 37:1609–1616.
- Hansen DJ, Parrish PR, Schimmel SC, Goodman LR. 1978. Sheepshead minnows (*Cyprinodon variegatus*). Bioassay procedures for the ocean disposal permit program. Washington DC: US Environmental Protection Agency, EPA-600/9:78-010.
- Hutchinson TH, Ankley GT, Segner H, Tyler CR. 2006. Screening and testing for endocrine disruption in fish-biomarkers as signposts not traffic lights in risk assessment. *Environ Health Perspect* 114 (Suppl 1):106–114.
- Jørgensen A, Morthorst JE, Andersen O, Rasmussen LJ, Bjerregaard P. 2008. Expression profiles for six zebrafish genes during gonadal sex differentiation. *Reprod Biol Endocrinol* 6:25.
- Kidd KA, Blanchfield PJ, Mills KH, Palace VP, Evans RE, Lazorchak JM, Flick RW. 2007. Collapse of a fish population after exposure to a synthetic estrogen. *Proc Natl Acad Sci U S A* 104:8897–8901.
- Kiker GA, Bridges TS, Varghese A, Seager TP, Linkov I. 2005. Application of multicriteria decision analysis in environmental decision making. *Integr Environ Assess Manag* 1:95–108.
- Länge R, Hutchinson TH, Croudace CP, Siegmund F. 2001. Effects of the synthetic estrogen 17 α -ethinylestradiol on the life cycle of the fathead minnow (*Pimephales promelas*). *Environ Toxicol Chem* 20:1216–1227.
- Lawrence C, Ebersole JP, Kesseli RV. 2008. Rapid growth and out-crossing promote female development in zebrafish (*Danio rerio*). *Environ Biol Fish* 81:239–246.
- Lorenzen K. 2005. Population dynamics and potential of fisheries stock enhancement practical theory for assessment and policy analysis. *Philos Trans R Soc Lond* 360:171–189.
- Lorenzen K. 2008. Fish population regulation beyond 'stock and recruitment': The role of density-dependent growth in the recruited stock. *Bull Mar Sci* 81:181–196.
- Martinović D, Hogarth WT, Jones RE, Sorensen PW. 2007. Environmental estrogens suppress hormones, behaviour, and reproductive fitness in male fathead minnows. *Environ Toxicol Chem* 26:271–278.
- Matsuda M, Nagahama Y, Shinomiya A, Sato T, Matsuda C, Kobayashi T, Morrey CE, Shibata N, Asakawa S, Shimizu N, Hori H, Hamaguchi S, Sakaizumi M. 2002. DMY is a Y-specific DM-domain gene required for male development in the medakafish. *Nature* 417:559–563.
- Matthiessen P. 2006. Estrogenic contamination of surface waters and its effects on fish in the United Kingdom. In: Vethaak D, Schrap M, de Voogt P, editors. Estrogens and xenestrogens in the aquatic environment: An integrated approach for field monitoring and effect assessment. Pensacola (FL): Society of Environmental Toxicology and Chemistry. p 329–357.
- Maunder RJ, Matthiessen P, Sumpter JP, Pottinger TG. 2007. Impaired reproduction in three-spined sticklebacks exposed to ethinyl estradiol as juveniles. *Biol Reprod* 77:999–1006.
- Miller DH, Jensen KM, Villeneuve DL, Kahl MD, Makynen EA, Durhan EJ, Ankley GT. 2007. Linkage of biochemical responses to population-level effects: A case study with vitellogenin in the fathead minnow (*Pimephales promelas*). *Environ Toxicol Chem* 26:521–527.
- Mills LJ, Chichester C. 2005. Review of evidence: Are endocrine-disrupting chemicals in the aquatic environment impacting fish populations? *Sci Total Environ* 343:1–34.
- Moore A, Lower N. 2001. The impact of two pesticides on olfactory-mediated endocrine function in mature male Atlantic salmon (*Salmo salar* L.) parr. *Comp Biochem Physiol B* 129:269–276.
- Nanda I, Kondo M, Hornung U, Asakawa S, Winkler C, Shimizu A, Shan Z, Haaf T, Shimizu N, Shima A, Schmid M, Schartl M. 2002. A duplicated copy of DMRT1 in the sex-determining region of the Y chromosome of the medaka. *Oryzias latipes*. *Proc Natl Acad Sci U S A* 99:11778–11783.
- Nash J, Kime D, Van der Ven L, Wester P, Brion F, Maack G, Stahlschmidt-Allner P, Tyler C. 2004. Long-term exposure to environmental concentrations of the pharmaceutical ethinylestradiol cause reproductive failure in fish. *Environ Health Perspect* 112: 1725–1733.
- OECD. 2002. A conceptual framework for the testing and assessment of endocrine disrupting chemicals. [Internet] Paris: Organisation for Economic Cooperation and Development; [cited 2009 March 26]. Available from: http://www.oecd.org/document/58/0,3343,en_2649_34377_2348794_1_1_1_1,00.html
- OECD. 2008. Draft detailed review paper on fish life cycle toxicity testing. ENV/JM/TG/EDTA(2008)2. Paris: Organisation for Economic Cooperation and Development. 154 p.
- Parrott JL, Blunt BR. 2005. Life cycle exposure of fathead minnows (*Pimephales promelas*) to an ethinylestradiol concentration below 1ng l⁻¹ reduces egg fertilization success and demasculinizes males. *Environ Toxicol* 20:131–141.
- Rose KA, Cowan JH, Winemiller KO, Myers RA, Hilborn R. 2001. Compensatory density-dependence in fish populations: Importance, controversy, understanding and prognosis. *Fish Fish* 2:293–327.
- Sparling GP, Tarbotton IS. 2000. Workshop to develop soil quality indicators. Lincoln (NZ): Landcare Research. Contract Report LC9900/078.
- Thorpe JE. 2004. Life history responses of fishes to culture. *J Fish Biol* 65 (Suppl 1): 263–285.
- [USEPA] US Environmental Protection Agency. 2002. Draft detailed review paper a fish two-generation toxicity test. EPA contract number 68-W-01-023.
- Van den Belt K, Verheyen R, Witters H. 2003. Effects of 17 α -ethinylestradiol in a partial life cycle test with zebrafish (*Danio rerio*): Effects of growth, gonads and female reproductive success. *Sci Total Environ* 309:127–137.
- Wenzel A, Schäfers C, Vollmer G, Michna H, Diel P. 2001. Research efforts towards the development and validation of a test method for the identification of endocrine disrupting chemicals. Fraunhofer Institute research report B6-7920/98/000015. 82 p.
- World Health Organisation. 2002. Global assessment of the state-of-the-science of endocrine disruptors. International Programme on Chemical Safety. WHO/PCS/EDC/02.2.
- Yatsalo BI, Kiker GA, Kim J, Bridges TS, Seager TP, Gardner K, Satterstrom FK, Linkov I. 2007. Application of multicriteria decision analysis tools to two contaminated sediment case studies. *Integr Environ Assess Manag* 3:223–233.
- Zha J, Sun L, Zhou Y, Spear PA, Ma M, Wang Z. 2008. Assessment of 17 α -ethinylestradiol effects and underlying mechanisms in a continuous, multigeneration exposure of the Chinese rare minnow (*Gobiocypris rarus*). *Toxicol Appl Pharmacol* 226:298–308.