OECD GUIDELINE FOR TESTING OF CHEMICALS

Adopted by the Council on 17th July 1992

Fish, Early-life Stage Toxicity Test

INTRODUCTION

1. Tests with the early-life stages of fish are intended to define the lethal and sub-lethal effects of chemicals on the stages and species tested. They yield information of value for the estimation of the chronic lethal and sub-lethal effects of the substance on other fish species.

2. This guideline is based on a proposal from the United Kingdom which was discussed at a meeting of OECD experts convened at Medmenham (United Kingdom) in November 1988 and further updated in 2012 to reflect experience in using the test and recommendations from an OECD workshop on fish toxicity testing, held in September 2010 (1).

PRINCIPLE OF THE TEST

3. The early-life stages of fish are exposed to a range of concentrations of the test substance dissolved in water. Flow through conditions are preferred, however if it is not appropriate semi-static conditions are acceptable. For details the OECD Guidance Document on aquatic toxicity testing of difficult substances and mixtures should be consulted (2). The test is initiated by placing fertilised eggs in test chambers and is continued until a juvenile lifestage, that is a species specific number of days post hatch. Lethal and sub-lethal effects are assessed and compared with control values to determine the lowest observed effect concentration in order to determine the (i) no observed effect concentration and/or (ii) ECx (e.g. EC10, EC_{20}) by using a regression model to estimate the concentration that would cause a x % change in the effect measured. The test concentrations should bracket the EC_x (e.g. EC_{10}, EC_{20}, EC_{50}) so that the EC_x comes from interpolation rather than extrapolation (see Annex 1 for definitions).

INFORMATION ON THE TEST SUBSTANCE

4. Results of an acute toxicity test (see Guideline 203), preferably performed with the species chosen for this test, should be available. This implies that the water solubility and the vapour pressure of the test substance are known and a reliable analytical method for the quantification of the substance in the test solutions with known and reported accuracy and limit of detection is available.

5. Useful information includes the structural formula, purity of the substance, stability in water and light, pK_a, P_{ow} and results of a test for ready biodegradability (see Guideline 301).

VALIDITY OF THE TEST

6. For a test to be valid the following conditions apply:

- the dissolved oxygen concentration should be >60% of the air saturation value throughout the test;

- the water temperature should not differ by more than $\pm$ 1.5°C between test chambers or between successive days at any time during the test, and should be within the temperature

ranges specified for the test species (Annex 3);

- evidence should be available to demonstrate that the concentrations of the test substance in solution have been satisfactorily maintained within ± 20% of the mean measured values;

- overall survival of fertilised eggs in the controls and, where relevant, in the solvent-only controls should be greater than or equal to the limits defined in Annex 3 ;

- when a solubilising agent is used it should have no significant effect on survival nor produce adverse effects on hatch or growth as revealed by a solvent-only control.

If a minor deviation from the test acceptance criteria is observed, the consequences should be considered in relation to the reliability of the test data and these considerations should be included in the report.

DESCRIPTION OF THE METHOD

Test chambers

7. Any glass, stainless steel or other chemically inert vessels can be used. The dimensions of the vessels should be large enough to allow proper growth in the control, maintenance of dissolved oxygen concentration (e.g for small fish species, a 7 L tank volume enables to achieve this) and compliance with loading rate criteria given below. It is desirable that test chambers be randomly positioned in the test area. A randomised block design with each treatment being present in each block is preferable to a completely randomised design. The test chambers should be shielded from unwanted disturbance.

Selection of species

8. Recommended fish species are given in Table 1. This does not preclude the use of other species, but the test procedure may have to be adapted to provide suitable test conditions. The rationale for the selection of the species and the experimental method should be reported in this case.

Holding of the brood fish

9. Details on holding the brood stock under satisfactory conditions may be found in Annex 2 and the references cited (3) (4) (5).

Handling of embryos and larvae

10. Initially, embryos and larvae may be exposed within the main vessel in smaller glass or stainless steel vessels, fitted with mesh sides or ends to permit a flow of test solution through the vessel. Non-turbulent flow through these small vessels may be induced by suspending them from an arm arranged to move the vessel up and down but always keeping the organisms submerged. Fertilised eggs of salmonid fishes can be supported on racks or meshes with apertures sufficiently large to allow larvae to drop through after hatching.

11. Where egg containers, grids or meshes have been used to hold eggs within the main test vessel, these restraints should be removed after the larvae hatch, according to the guidance in Annex 2, except that meshes should be retained to prevent the escape of the fish. If there is a need to transfer the larvae, they should not be exposed to the air and nets should not be used to release fish from egg containers. The timing of this transfer varies with the species and transfer may not always be necessary.

Water

12. Any water in which the test species shows control survival and growth at least as good as that described in Annex 3 is suitable as a test water. It should be of constant quality during the period of the test. In order to ensure that the dilution water will not unduly influence the test result (for example by complexation of test substance) or adversely affect the performance of the brood stock, samples should be taken at intervals for analysis. Measurements of heavy metals (e.g. Cu, Pb, Zn, Hg, Cd, Ni), major anions and cations (e.g. Ca, Mg, Na, K, Cl, $S0_4$), ammonia, total residual chlorine pesticides, total organic carbon and suspended solids should be made, for example on a annual basis where a dilution water is known to be relatively constant in quality. Some chemical characteristics of an acceptable dilution water are listed in Annex 4.

Test solutions

13. For flow-through tests, a system which continually dispenses and dilutes a stock solution of the test substance (e.g. metering pump, proportional diluter, saturator system) is required to deliver a series of concentrations to the test chambers. The flow rates of stock solutions and dilution water should be checked at intervals during the test and should not vary by more than 10% throughout the test. A flow rate equivalent to at least five test chamber volumes per 24 hours has been found suitable (3). However if the loading rate specified in paragrapgh 17 is respected, a lower flowrate of e.g. 2-3 test chamber volume is possible to prevent quick removal of food.

14. Test solutions of the chosen concentrations are prepared by dilution of a stock solution. The stock solution should preferably be prepared by simply mixing or agitating the test substance in dilution water by using mechanical means (e.g. stirring or ultrasonication). Saturation columns (solubility columns) or passive dosing methods (6) can be used for achieving a suitable concentrated stock solution. The use of a solvent carrier is not recommended. However, in case a solvent is necessary, a solvent control should be run in parallel, at the same solvent concentration as the chemical treatments; i. e. the solvent level should be equal across all concentrations as well as the solvent control. For difficult to test substances, a solvent may be technically the best solution; the OECD Guidance Document on aquatic toxicity testing of difficult substances and mixtures should be consulted (2). The choice of solvent will be determined by the chemical properties of the substance. The OECD Guidance Document recommends a maximum concentration of 100µl/L. Further, a recent review observed evidence that some low concentrations of solvents may affect the reproduction of certain fish species, and also impact biomarkers of endocrine disruption (7). Therefore, where ever practically feasible, the use of solvents should be avoided.

15. For a semi-static technique, two different renewal procedures may be followed. Either new test solutions are prepared in clean vessels and surviving eggs and larvae gently transferred into the new vessels, or the test organisms are retained in the test vessels whilst a proportion (at least two thirds) of the test solution / control might is changed.

PROCEDURE

Conditions of Exposure

Duration

16. The test should start as soon as possible after the eggs have been fertilised, the embryos preferably being immersed in the test solutions before cleavage of the blastodisc commences, or as close as possible after this stage. The test duration will depend upon the species used. Some recommended durations are given in Annex 3 .

Loading

17. The number of fertilised eggs at the start of the test should be sufficient to meet statistical requirements. They should be randomly distributed among treatments, and at least 80 eggs, divided equally between at least four replicate test chambers, should be used per concentration. The loading rate (biomass per volume of test solution) should be low enough in order that a dissolved oxygen concentration of at least 60% of the air saturation value (ASV) can be maintained without aeration during the egg and larval stage. For flow-through tests, a loading rate not exceeding 0.5 g/L per 24 hours and not exceeding 5 g/L of solution at any time has been recommended (3).

Light and temperature

18. The photoperiod and water temperature should be appropriate for the test species (see Annex 3).

Feeding

19. Food and feeding are critical, and it is essential that the correct food for each stage should be supplied from an appropriate time and at a level sufficient to support normal growth. Feeding should be approximately equal across replicates unless adjusted to account for mortality. Surplus food and faeces should be removed as necessary to avoid accumulation of waste. Detailed feeding regimes are given in Annex 2 but, as experience is gained, food and feeding regimes are continually being refined to improve survival and optimise growth.

Test concentrations

20. Normally five concentrations of the test substance spaced by a constant factor not exceeding 3.2 are required. Information of the acute testing, preferable with the same species and/or a range finding test should be considered (1) when selecting the range of test concentrations. A limit test or extended limit test with fewer than five concentrations as the definitive test may be acceptable where empirical NOECs are to be established. Justification should be provided if fewer than five concentrations are used. Concentrations of the substance higher than the 96 hour LC_{50} or 10 mg/l, whichever is the lower, need not be tested

Controls

21. A dilution-water control and also, if relevant, a solvent control containing the solvent carrier only should be run in addition to the test substance concentration series.

Frequency of Analytical Determinations and Measurements

22. Prior to initiation of the exposure period, proper function of the chemical delivery system across all replicates should be ensured (e.g. for example by measuring test concentrations). All analytical methods needed should be established, including sufficient knowledge on the substance stability in the test system. During the test, the concentrations of the test substance are determined at regular intervals to characterise exposure. A minimum of five determinations is necessary. In flow-through systems, analytical measurements of the test chemical in one replicate per concentration should be made at least once a week changing systematically between replicates. Additional analytical determinations will often improve quality of the test outcome. Samples may need to be filtered to remove any particulate matter (e.g. using a 0.45 μm pore size) or centrifuged to ensure that the determinations are made on the substance in true solution. For the analysis a suitable analytical method is required with an appropriate limit of quantification (LOQ).

23. During the test, dissolved oxygen, pH, and temperature should be measured in all test vessels, at least weekly, and salinity and hardness, if warranted, at the beginning and end of the test. Temperature should preferably be monitored continuously in at least one test vessel.

Observations

24. **Stage of embryonic development:** the embryonic stage at the beginning of exposure to the test substance should be verified as precisely as possible. This can be done using a representative sample of eggs suitably preserved and cleared.

25. **Hatching and survival:** observations on hatching and survival should be made at least once daily and numbers recorded. Dead embryos, larvae and juvenile fish should be removed as soon as observed since they can decompose rapidly and may be broken up by the actions of the other fish. Extreme care should be taken when removing dead individuals not to knock or physically damage adjacent eggs/larvae, these are extremely delicate and sensitive. Signs of death vary according to species and life stage. For example:

- for fertilized eggs: particularly in the early stages, a marked loss of translucency and change in colouration, caused by coagulation and/or precipitation of protein, leading to a white opaque appearance;

- for embryos, larvae and juvenile fish: immobility and/or absence of respiratory movement and/or absence of heart-beat and/or lack of reaction to mechanical stimulus.

26. **Abnormal appearance:** the number of larvae or fish showing abnormality of body form should be recorded at adequate intervals depending on the duration of the test and the nature of the abnormality described. It should be noted that abnormal embryos and larvae occur naturally and can be of the order of several per cent in the control(s) in some species. Where deformities and associated abnormal behaviour are considered so severe that there is considerable suffering to the organism and it has reached a point beyond which it will not recover it may be removed from the test. Such animals should be humanely euthanized and treated as mortalities for subsequent data analysis. Normal embryonic development has been documented for most species recommended in this Guideline (8) (9) (10) (11).

27. **Abnormal behaviour:** abnormalities, e.g. hyperventilation, uncoordinated swimming, atypical quiescence and atypical feeding behaviour should be recorded at adequate intervals depending on the duration of the test. These effects, although difficult to quantify, can, when observed, aid in the interpretation of mortality data.

28. **Weight:** at the end of the test, all surviving fish are weighed on a replicate basis (reporting the number of animals in the replicate and the mean weight per animal): wet weight – (blotted dry) is preferred, additionally dry weight data may also be reported (12).

29. **Length:** at the end of the test, individual lengths are measured. Total length is recommended, if however, caudal fin rot or fin erosion occurs, standard lengths can be used. Individual length can be measured either by callipers, digital camera, or calibrated ocular micrometer. Typical minimal lengths are defined in Annex 3..

30. These observations will result in some or all of the following data being available for statistical analysis:

- cumulative mortality;
- numbers of healthy fish at end of test;
- time to start of hatching and end of hatching;
- numbers of larvae hatched;
- length and weight of surviving animals;
- numbers of deformed larvae;
- numbers of fish exhibiting abnormal behaviour.

DATA AND REPORTING

Treatment of results

31. The design of the experiment and selection of statistical test should permit adequate power (75% or higher) to detect changes of biological importance in the more important endpoints where a NOEC is to be reported. If an ECx is to be reported, the design of the experiment and selection of regression model should permit estimation of ECx so that (i) the 95% confidence interval reported for ECx does not contain zero and is not overly wide, (ii) the 95% confidence interval predicted mean at ECx does not contain the control mean (iii) there is no significant lack-of-fit of regression model to the data. Either approach requires the identification of the percent change in each endpoint that is important to detect or estimate. The experimental design should be tailored to allow that. It is not likely that the same percent change applies to all endpoints, nor is it likely that a feasible experiment can be designed that will meet these criteria for all endpoints, so it is important to focus on the most important endpoints in designing the experiment. Statistical flow diagrammes and guidance for each approach are available in Annexes 5 and 6 to guide in the treatment of data and in the choice of the most appropriate statistical test or model to use.

32. It will be necessary for variations to be analysed within each set of replicates using analysis of variance or contingency table procedures and appropriate statistical analysis methods be used based on this analysis. In order to make a multiple comparison between the results at the individual concentrations and those for the controls, the step-down Jonckheere-Terpstra or Williams' test is recommended for continuous responses and a step-down Cochran-Armitage test for quantal responses that are consistent with a monotone concentration-response and with no evidence of extra-binomial variance (13). When there is evidence of extra-binomial variance, the Rao-Scott modification of the Cochran-Armitage test is recommended (14) (15) or Williams or Dunnett's (after an arcsin-square-root transform) or Jonckheere-Terpstra test applied to replicate proportions. Where the data are not consistent with a monotone concentration-response, Dunnett's or Dunn's or the Mann-Whitney method may be found useful for continuous responses and Fisher's Exact test for quantal responses (13) (16) (17). Care must be taken where applying any statistical method or model to ensure that the requirements of the method or model are satisfied (e.g. chamber to chamber variability is estimated and accounted for in the experimental design and test or model used). Transformations to meet the requirements of a statistical test should be considered. However, transformations to enable the fitting of a regression model require great care, as, for example, a 25% change in the untransformed response does not correspond to a 25% change in a transformed response. In all analyses, the test chamber, not the individual fish, is the unit of analysis and the experimental unit and both hypothesis tests and regression should reflect that (3) (13) (18) (19).

Test report

33. The test report must include the following information:

Test substance:
- physical appeareance and, where relevant, physicochemical properties;
- chemical identification data.

Test species:
- scientific name, strain, source and method of collection of the fertilised eggs and subsequent handling.

Test conditions:
- test procedure used (e.g. semi-static or flow-through, loading);
- photoperiod(s);
- test design (e.g. number of test chambers and replicates, number of embryos per replicate, size of the test chamber, water volume per test chamber);
- method of preparation of stock solutions and frequency of renewal (the solubilising

agent and its concentration must be given, when used);
- method of dosing the test substance (e.g. pumps, diluting systems)
- the nominal test concentrations, the means of the measured values and their standard deviations in the test vessels and the method by which these were attained and evidence that the measurements refer to the concentrations of the test substance in true solution;
- dilution water characteristics: pH, hardness, temperature, dissolved oxygen concentration, residual chlorine levels (if measured), total organic carbon (if measured), suspended solids (if measured), salinity of the test medium (if measured) and any other measurements made;
- water quality within test vessels, pH, hardness, temperature and dissolved oxygen concentration;
- detailed information on feeding (e.g. type of food(s), source, amount given and frequency).

Results reported individually (or on a replicate basis) and as mean and coefficient of variation, as appropriate, for the following endpoints:
- evidence that controls met the overall survival acceptability standard of the test species (Annex 3);
- data on mortality at each stage (embryo, larval and juvenile) and overall mortality;
- days to hatch and numbers hatched;
- data for length (specify either standard or total) and weight;
- incidence and description of morphological abnormalities, if any;
- incidence and description of behavioural effects, if any;
- approach for the statistical analysis (regression analysis or analysis of the variance) and treatment of data (statistical test or model used);
- no observed effect concentration for each response assessed (NOEC);
- lowest observed effect concentration (at $p = 0.05$) for each response assessed (LOEC);
- EC_x for each response assessed and confidence intervals (e.g. 90% or 95%) and a graph of the fitted model used for its calculation, the slope of the concentration-response curve and its standard error;
- any concentration-response data and curves available.

Discussion of the results, including any influence on the outcome of the test resulting from deviations from this Guideline.

TABLE 1: FISH SPECIES RECOMMENDED FOR TESTING

FRESHWATER	BRACKISH WATER	SALTWATER
Oncorhynchus mykiss Rainbow trout **Pimephales promelas** Fathead minnow **Danio rerio** Zebrafish **Oryzias latipes** Japanese ricefish or Medaka	**Cyprinodon variegatus** Sheepshead minnow	**Menidia sp.** Silverside

LITERATURE

(1) OECD (2012). Fish Toxicity Testing Framework. OECD Environmental Health and Safety Publications Series on Testing and Assessment No.171. OECD, Paris

(2) OECD (2000). Guidance Document on Aquatic Toxicity Testing of Difficult Substances and Mixtures. Environmental Health and Safety Publications. Series on Testing and Assessment. No. 23. Paris

(3) ASTM (1988). Standard Guide for Conducting Early Life-Stage Toxicity Tests with Fishes. American Society for Testing and Materials. E 1241-88. 26 pp.

(4) Brauhn J.L. and Schoettger R.A. (1975). Acquisition and Culture of Research Fish: Rainbow trout, Fathead minnows, Channel catfish and Bluegills. p. 54, Ecological Research Series, EPA-660/3-75-011, Duluth, Minnesota.

(5) Brungs W.A. and Jones B.R. (1977). Temperature Criteria for Freshwater Fish: Protocol and Procedures. p. 128, Ecological Research Series EPA-600/3-77-061, Duluth, Minnesota.

(6) Adolfsson-Erici, M., Åkerman, G., Jahnke, A., Mayer, P., McLachlan, M.S. (2012). A flow-through passive dosing system for continuously supplying aqueous solutions of hydrophobic chemicals to bioconcentration and aquatic toxicity tests. Chemosphere 86, 593-599.

(7) Hutchinson, T.H.; Shillabeer, N., Winter, M.J. and Pickford, D.B. (2006). Acute and chronic effects of carrier solvents in aquatic organisms: A critical review. Aquatic Toxicology, 76, 69-92

(8) Hansen D.J. and Parrish P.R. (1977). Suitability of sheepshead minnows (*Cyprindon variegatus*) for life-cycle toxicity tests. In Aquatic Toxicology and Hazard Evaluation (edited by F.L. Mayer and J.L. Hamelink), pp. 117-126, ASTM STP 634.

(9) Kimmel, H. B.; Ballard, W. W.; Kimmel, S. R.; Ullman, B.; Schilling, T. F. (1995). Stages of embryonic development of the zebrafish. Dev. Dyn. 203:253–310.

(10) Gonzalez-Doncel, M.; Okihiro, M.S.; Villalobos, S.A.; Hinton D.E.; Tarazona J.V. (2005). A quick reference guide to the normal development of *Oryzias latipes* (Teleostei, Adrinichthydae) J. Appl. Ichthyol., 20:1–14

(11) Devlin EW, Brammer JD, Puyear RL, McKim JM. (1996) Prehatching Development of the Fathead Minnow, *Pimephales promelas* Rafinesque. EPA/600/R-96/079. USEPA, Office of Research and Development, Washington, D.C. 57p.

(12) Oris, J.T.; Belanger, S.C. and Bailer, A.J. (2012). Baseline characteristics and statistical implications for the OECD 210 Fish Early Life Stage Chronic Toxicity Test. Environmental Toxicology and Chemistry 31; 2, 370 – 376.

(13) OECD (2006). Current Approaches in the Statistical Analysis of Ecotoxicity Data: A Guidance to Application. OECD environmental Health and Safety Publications Series on Testing and Assessment No.54. ENV/JM/MONO(2006)18

(14) Rao J.N.K. and Scott A.J. (1992). A simple method for the analysis of clustered binary data. Biometrics 48, 577-585.

(15) Rao J.N.K. and Scott A.J. (1999). A simple method for analyzing overdispersion in clustered Poisson data. Statistics in Medicine 18, 1373-1385.

(16) Dunnett C.W. (1955). A multiple comparisons procedure for comparing several treatments with a control. J. Amer, Statist. Assoc., 50, 1096-1121.

(17) Dunnett C.W. (1964). New tables for multiple comparisons with a control. Biometrics, 20, 482-491.

(18) Rand G.M. and Petrocelli S.R. (1985). Fundamentals of Aquatic Toxicology. Hemisphere Publication Corporation, New York.

(19) McClave J.T., Sullivan J.H. and Pearson J.G. (1980). Statistical Analysis of Fish Chronic Toxicity Test Data. Proceedings of 4th Aquatic Toxicology Symposium, ASTM, Philadelphia.

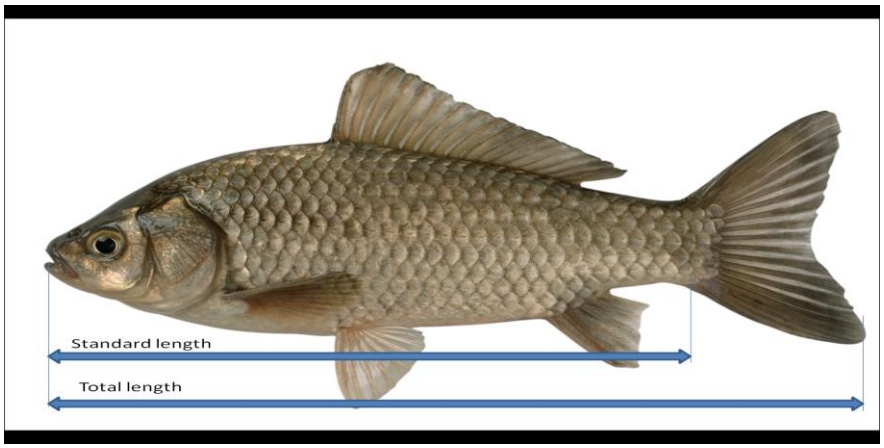
ANNEX 1

DEFINITIONS

Fork length (FL): refers to the length from the tip of the snout to the end of the middle caudal fin rays and is used in fishes in which it is difficult to tell where the vertebral column ends (www.fishbase.org)

Standard length (SL): refers to the length of a fish measured from the tip of the snout to the posterior end of the last vertebra or to the posterior end of the midlateral portion of the hypural plate. Simply put, this measurement excludes the length of the caudal fin. (www.fishbase.org)

Total length (TL): refers to the length from the tip of the snout to the tip of the longer lobe of the caudal fin, usually measured with the lobes compressed along the midline. It is a straight-line measure, not measured over the curve of the body (www.fishbase.org)



ECx: (Effect concentration for x% effect) is the concentration that causes an x% of an effect on test organisms within a given exposure period when compared with a control. For example, an EC50 is a concentration estimated to cause an effect on a test end point in 50% of an exposed population over a defined exposure period.

Lowest observed effect concentration (LOEC) is the lowest tested concentration of a test substance at which the substance is observed to have a significant effect (at $p < 0.05$) when compared with the control. However, all test concentrations above the LOEC must have a harmful effect equal to or greater than those observed at the LOEC Annexes 5 and 6 provide guidance.

No observed effect concentration (NOEC) is the test concentration immediately below the LOEC.

ANNEX 2

FEEDING AND HANDLING GUIDANCE FOR BROOD AND TEST ANIMALS OF RECOMMENDED SPECIES

SPECIES	FOOD					POST-HATCH TRANSFER TIME	TIME TO FIRST FEEDING
	Brood fish	Newly-hatched larvae	Juveniles				
			Type	Amount	Frequency		
Freshwater:							
Oncorhynchus mykiss Rainbow trout	trout food	None [a]	trout starter	To satiation	2-4 feeds per day	14-16 days post-hatch or at swim-up (not essential)	19 days post-hatch or at swim-up
Pimephales promelas Fathead minnow	FBS	BSN	BSN48, flake food	To satiation	2-3 times a day	once hatching is 90%	2 day post hatch
Danio rerio Zebrafish	BSN48, flake food	Commercial larvae food protozoa[b], protein[c]	BSN48, at least once daily flake food, at least twice daily	To satiation	BSN once daily; flake food twice daily	once hatching is 90%	2 days post hatch
Oryzias latipes Japanese Ricefish or Medaka	flake food	BSN, flake food (or protozoa or rotifers)	BSN48, flake food (or rotifers)	To satiation	BSN once daily; flake food twice daily or flake food and rotifers once daily	not applicable	6-7 days post spawn
Brackish water:							
Cyprinodon variegatus Sheepshead minnow	FBS or flake food	BSN	BSN48	To satiation	2-3 feeds per day	not applicable	1 day post hatch/swim-up
Salt water:							
Menidia sp. Silverside	BSN48, flake food	BSN	BSN48	To satiation	2-3 feeds per day	not applicable	1 day post hatch/swim-up

Key:

FBS frozen brine shrimps; adults Artemia sp
BSN brine shrimp nauplii; newly hatched
BSN48 brine shrimp nauplii; 48 hours old

(a) yolk-sac larvae require no food
(b) filtered from mixed culture
(c) granules from fermentation process

ANNEX 3

TEST CONDITIONS, DURATION AND SURVIVAL CRITERIA FOR RECOMMENDED SPECIES

SPECIES	TEST CONDITIONS			RECOMMENDED DURATION OF TEST	Typical minimum mean total length of control fish at the end of the study (mm)[1]	SURVIVAL OF CONTROLS (minimum)	
	Temperature (°C)	Salinity ($^0/_{00}$)	Photoperiod (hrs)			Hatching success	Post-hatch success
Freshwater:							
Oncorhynchus mykiss Rainbow trout	10 ± 1.5[2]		(3)	2 weeks after controls are free-feeding (or 60 days post-hatch)	40	75%	75%
Pimephales promelas Fathead minnow	25 ± 1.5		16	32 days from start of test (or 28 days post-hatch)	18[6]	70%	75%
Danio rerio Zebrafish	26 ± 1.5		12 - 16[4]	30 days post-hatch	11[6]	70%	75 %
Oryzias latipes Japanese Medaka or Medaka	25 ± 2		12 - 16[4]	30 days post-hatch	17	80%	80%
Brackish water:							
Cyprinodon variegatus Sheepshead minnow	25 ± 1.5	15 - 30[5]	12 - 16[4]	32 days from start of test (or 28 days post-hatch)	17	75%	80%
Saltwater:							
Menidia sp. Silverside	22 - 25	15-30[5]	13	28 days	20	80%	60%

Key:

(1) Typical minimum mean total length is not a validity criterion but deviations below the figure indicated should be carefully examined in relation to the sensitivity of the test.

(2) The particular strain of rainbow trout tested may necessitate the use of other temperatures. Brood stock must be held at the same temperature as that to be used for the eggs.

(3) Darkness for larvae until one week after hatching except when they are being inspected, then subdued lighting throughout test (12-16 hour photoperiod).

(4) For any given test conditions, light regime should be constant.

(5) For any given test this shall be performed to $\pm 2^0/_{00}$.

(6) Oris et al. (2012)

ANNEX 4

SOME CHEMICAL CHARACTERISTICS OF AN ACCEPTABLE DILUTION WATER

SUBSTANCE	CONCENTRATIONS
Particular matter	< 20 mg/l
Total organic carbon	< 2 mg/l
Unionised ammonia	< 1 ug/l
Residual chlorine	< 10 ug/l
Total organophosphorus pesticides	< 50 ng/l
Total organochlorine pesticides plus polychlorinated biphenyls	< 50 ng/l
Total organic chlorine	< 25 ng/l

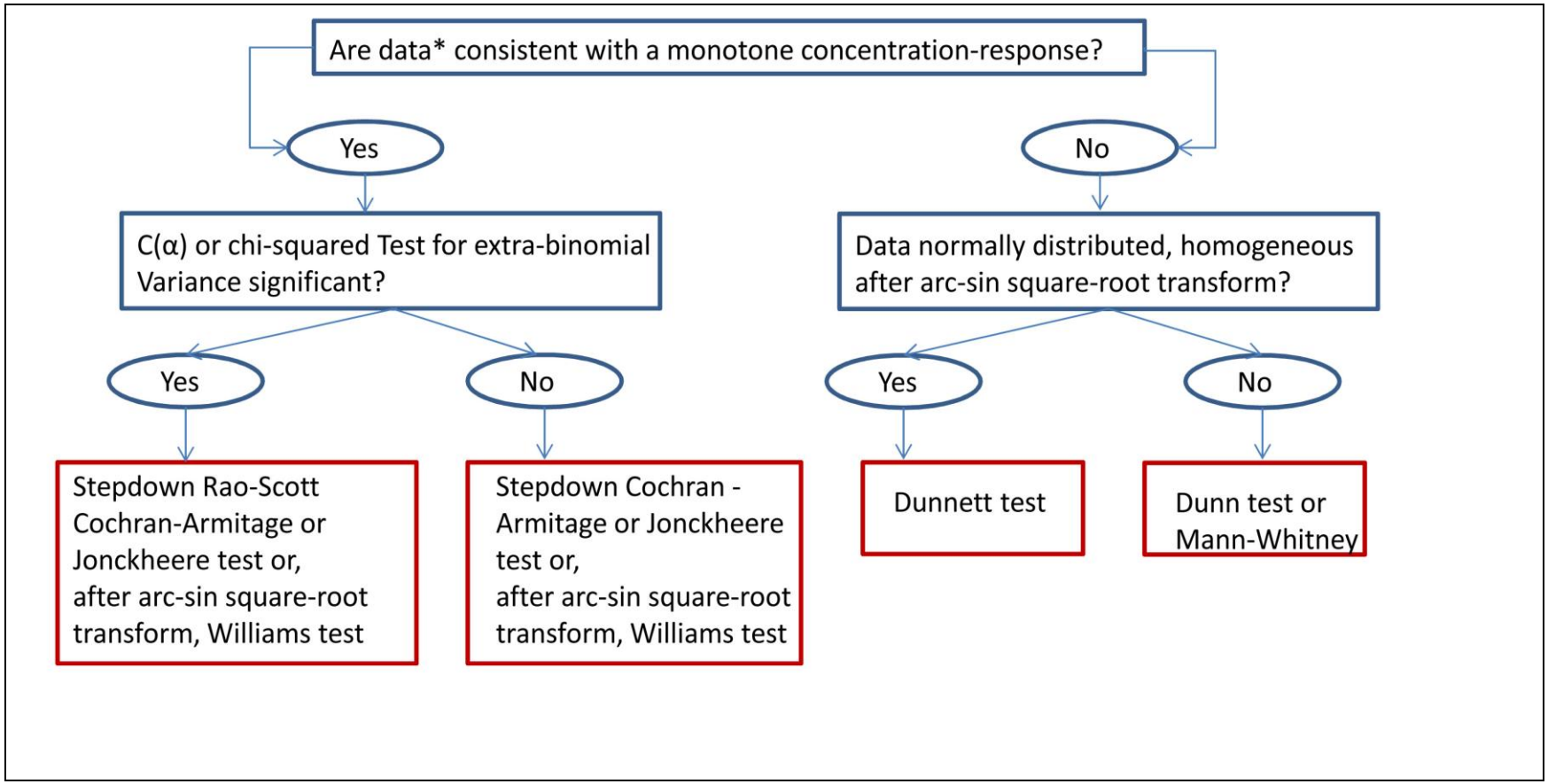


Figure 2: Statistical Flow Chart for Egg Hatch and Larval Mortality NOEC:
*Data are replicate proportion

ANNEX 5 Statistical Protocols for NOEC Determination (continued)

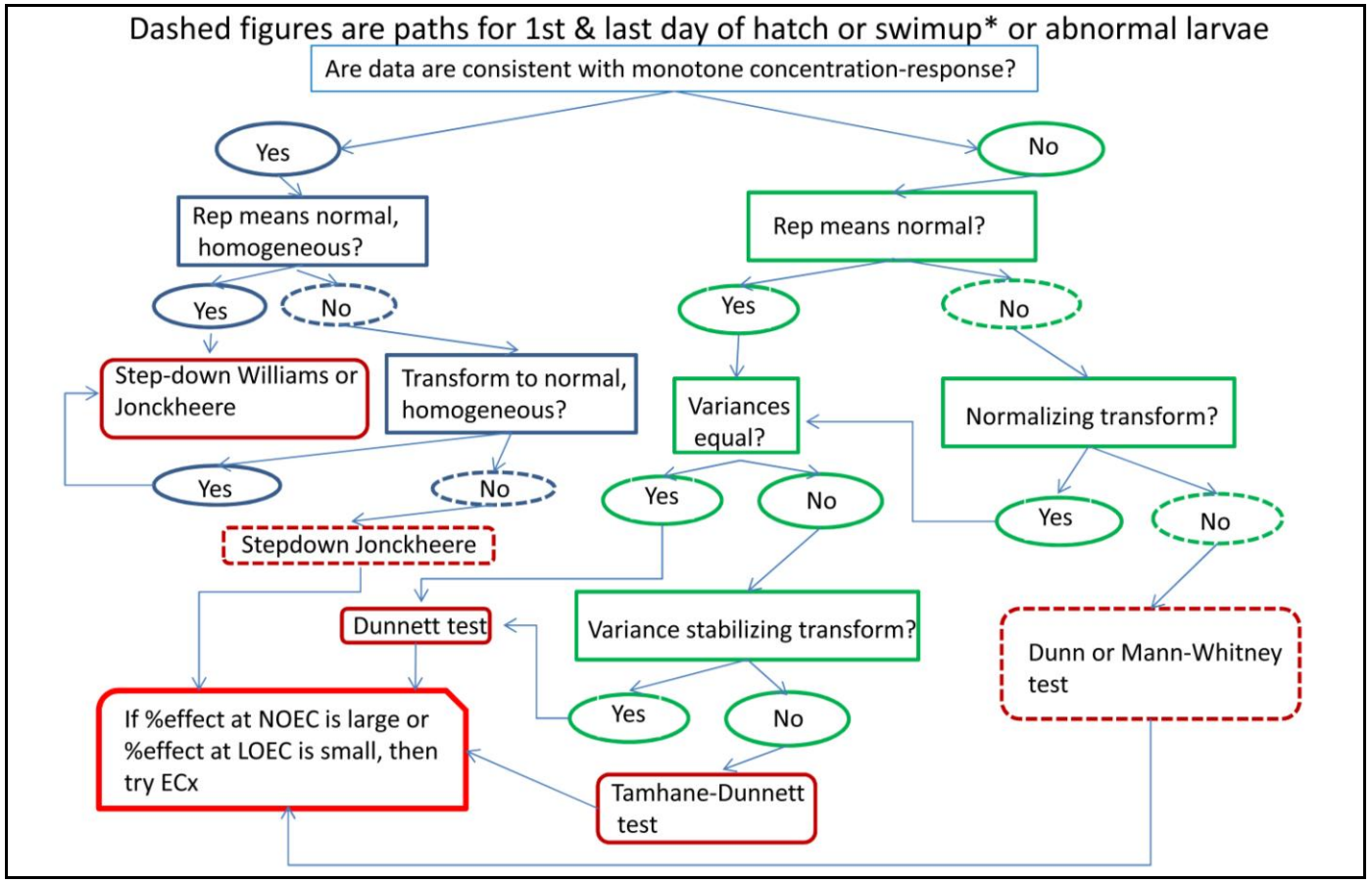


Figure 3: NOEC Flow-Chart for Length and Weight

*These responses never satisfy assumptions for parametric analysis or models

General

The replicate tank is the unit of analysis. Thus, for continuous measurements, such as size, the replicate mean or median should be calculated and these replicate values are the data for analysis. The power of the tests used should be demonstrated, preferably based on an adequate historical database for each lab. The size effect that can be detected with 75-80% power should be provided for each endpoint with the statistical test to be used.

- The databases available at the time of development of this guideline establish the power possible under the recommended statistical procedures. An individual lab should demonstrate its ability to meet this power requirement either by conducting its own power analysis or by demonstrating that the CV for each response does not exceed the 90th percentile of CVs used in developing the TG. Table 1 provides these CVs. If only replicate means or medians are available, then the within-rep CV can be ignored.

Table 1: **90th Percentile CVs for selected Freshwater Species**

Species	Response	CV_Between Replicates	CV_Within Replicates
Rainbow Trout	Length	17.4	9.8
	Weight	10.1	28
Fathead Minnow	Length	16.9	13.5
	Weight	11.7	38.7
Zebrafish	Length	43.7	11.7
	Weight	11.9	32.8

For almost all statistical tests used to evaluate laboratory toxicology studies, the comparisons of interest are of treatment groups to control. For that reason, it is not appropriate to require a significant ANOVA F-test before using Dunnett's or Williams' test or a significant Kruskal-Wallis test before using the Jonckheere-Terpstra, Mann-Whitney, or Dunn test (Hochberg and Tamhane 1987, Hsu 1996, Dunnett 1955, 1964, Williams 1971, 1972, 1975, 1977, Robertson et al. 1988, Jonckheere 1954, Dunn 1964).

Dunnett's test has a built-in multiplicity adjustment and its false positive and false negative rates are adversely affected by using the F-test as a gatekeeper. Similarly, the step-down Williams and Jonckheere-Terpstra tests using a 0.05 significance level at every step preserve an overall 5% false positive rate and that rate and the power of the tests are adversely affected by using the F- or Kruskal-Wallis test as a gatekeeper. Mann-Whitney and Dunn's test have to be adjusted for multiplicity and the Bonferroni-Holm adjustment is advised.

A thorough discussion of most of the recommendations on hypothesis testing and verification of assumptions underlying these tests is given in OECD (2006), which also contains an extensive bibliography.

Treatment of Controls when a Solvent is Used

If a solvent is used, then both water and a solvent controls should be included. The two controls should be compared for each response and combined for statistical analysis if no significant difference is found between the controls. Otherwise, the solvent control should be used for NOEC determination or ECx estimation and the water control is not used. See restriction in the validity criteria (Paragraph 6)
For length, weight, proportion of egg hatch or larval mortality or abnormal larvae, and first or last day of hatch or swimup, a T-test or Mann-Whitney test should be used to compare the controls at the significance at the 0.01 level, ignoring all treatment groups. The results of these tests should be reported.

Size Measurements (length and weight)

Individual fish length and weight values can be normally or log-normally distributed. In either case, the replicate mean values tend to be normally distributed by virtue of the Central Limit Theorem and confirmed from data from well over 100 ELS studies of three freshwater species. Alternatively, where the data or historical databases suggest a log-normal distribution for individual fish size values, the replicate mean logarithm of the individual fish values can be calculated and the data for analysis can then be the anti-logs of these replicate mean logarithms.

Data should be evaluated for consistency with a normal distribution and variance homogeneity. For this purpose, the residuals from an ANOVA model with concentration as the single explanatory class variable should be used. Visual determination from scatterplots and histograms or stem-and-leaf plots can be used. Alternatively, a formal test such as the Shapiro-Wilk or Anderson-Darling can be used. Consistency with variance homogeneity can be assessed from a visual examination of the same scatter plot or formally from Levene's test. Only parametric tests (e.g., Williams, Dunnett) need be evaluated for normality or variance homogeneity.

Attention should be paid to possible outliers and their effect on analysis. Tukey's outlier test and visual inspection of the same plots of residuals described above can be used. It should be recalled that observations are entire replicates, so omitting an outlier from analysis should be done only after careful consideration.

The statistical tests that make use of the characteristics of the experimental design and biological expectation are step-down trend tests, such as Williams and Jonckheere-Terpstra. These tests assume a monotone concentration-response and the data should be assessed for consistency with that assumption. This can be done visually from a scatter plot of the replicate means against test concentration. It will be helpful to overlay that scatter plot with a piecewise linear plot connecting the concentration means weighted by replicate sample size. Great deviation of this piecewise linear plot from monotonicity would indicate a possible need to use non-trend tests. Alternatively, formal tests can be used. A simple formal test is to compute linear and quadratic contrasts of the concentration means. If the quadratic contrast is significant and the linear contrast is not significant, that is an indication of a possible problem with monotonicity which should be further evaluated from plots. Where normality or variance homogeneity may be an issue, these contrasts can be constructed from rank-order transformed data. Alternative procedures, such as Bartholomew's test for monotonicity can be used, but add complexity.

Unless the data are not consistent with the requirements for these tests, the NOEC is determined by a step-down application of Williams' or the Jonckheere-Terpstra test. OECD (2006) provides details on these procedures. For data not consistent with the requirements for a step-down trend test, Dunnett's test or the Tamhane-Dunnett (T3) test can be used, both of which have built-in adjustments for multiplicity. These tests assume normality and, in the case of Dunnett, variance homogeneity. Where those conditions are not satisfied, Dunn's non-parametric test can be used. OECD (2006) contains details for all of these tests.

Egg Hatch and Larval Survival

The data are replicate proportions of eggs that hatch or larvae that survive. These proportions should be assessed for extra-binomial variance, which is common but not universal for such measurements. Two tests are commonly used. These are Tarone's C(α) test (Tarone, 1979) and chi-squared tests, each applied separately to every test concentration. If extra-binomial variance is found in even one test concentration, then methods that accommodate that should be used.

$$Z = \frac{\sum_{j=1}^{m} \frac{(x_j - n_j\hat{p})^2}{\hat{p}(1-\hat{p})} - \sum_{j=1}^{m} n_j}{\left\{2\sum_{j=1}^{m} n_j(n_j-1)\right\}^{1/2}}$$

Formula 1: Tarone's C (α) test (Tarone 1979)

where $\underline{\hat{p}}$ is the mean proportion for a given concentration, **m** is the number of replicate tanks, $\mathbf{n_j}$ is the number of subjects in replicate **j**, and $\mathbf{x_j}$ is the number of subjects in that replicate responding, e.g., not hatched or dead. This test is applied to each concentration separately. This test can be seen as an adjusted chi-squared test, but limited power simulations done by Tarone have shown it to be more powerful than a chi-squared test.

Where there is no significant evidence of extra-binomial variance, the step-down Cochran-Armitage test can be used. This test ignores replicates, so where there is such evidence, the Rao-Scott adjustment to the Cochran-Armitage test (RSCA) takes replicates, replicate sizes, and extra-binomial variance into account and is recommended. Alternative tests include the step-down Williams and Jonckheere-Terpstra tests and Dunnett's test as described for size measurements. These tests apply whether or not there is extra-binomial variance, but have somewhat lower power (Agresti 2002, Morgan 1992, Rao and Scott 1992, 1999, Fung et al. 1994, 1996)

First or Last Day of Hatch or Swimup

The response is an integer, giving the test day on which the indicated observation is observed for a given replicate tank. The range of values is generally very limited and there are often high proportions of tied values, e.g., the same first day of hatch is observed in all control replicates and, perhaps in one or two low test concentrations. Parametric tests such as Williams and Dunnett are not appropriate for such data. Unless there is evidence on serious non-monotonicity, the step-down Jonckheere-Terpstra test is very powerful for detecting effects of the test substance. Otherwise, Dunn's test can be used.

Larval Abnormalities

The response is the count of larvae found to be abnormal in some way. This response is frequently of low incidence and has some of the same problems as first day of hatch, as well as sometimes exhibiting erratic in concentration-response. If the data at least roughly follow a monotone concentration shape, the step-down Jonckheere-Terpstra test is powerful for detecting effects. Otherwise, Dunn's test can be used.

References

Agresti, A. (2002); Categorical Data Analysis, second edition, Wiley, Hoboken.

Berenson B.M. (1982a); A comparison of several k sample tests for ordered alternatives in completely randomized designs, Psychometrika 47, 265-280.

Berenson M. L. (1982b); A study of several useful tests for ordered alternatives in the randomized block design, Comm. Statistical (B) 11, 563-581.

Berenson M. L. (1982c); Some useful nonparametric tests for ordered alternatives in randomized block experiments, Comm. Statistical (A) 11, 1681-1693.

Dunn O. J. (1964); Multiple Comparisons Using Rank Sums, Technometrics 6, 241-252.

Dunnett C. W. (1964); New tables for multiple comparisons with a control, Biometrics 20, 482-491.

Dunnett C. W. (1955); A multiple comparison procedure for comparing several treatments with a control, J. American Statistical Association 50, 1096-1121.

Dunnett C. W. and Tamhane A. C. (1998); New multiple test procedures for dose finding, Journal of Biopharmaceutical Statistics, 8, 353 366.

Fung, K.Y., D. Krewski, J.N.K. Rao, A.J. Scott (1994); Tests for Trend in Developmental Toxicity Experiments with Correlated Binary Data, Risk Analysis 14, 639-648.

Fung, K.Y, D. Krewski, R.T. Smythe (1996); A comparison of tests for trend with historical controls in carcinogen bioassay, Canadian Journal of Statistics 24, 431-454.

Hochberg, Y. and A. C. Tamhane (1987); Multiple Comparison Procedures, Wiley, New York.

Hsu, J.C. (1996); Multiple Comparisons: Theory and Methods; Chapman and Hall/CRC Press, Boca Raton.

Jonckheere A. R. (1954); A distribution-free k-sample test against ordered alternatives, Biometrika 41, 133.

Kodell R.L. and Chen J.J. (1991); Characterization of Dose-Response Relationships Inferred by Statistically Significant Trend Tests, Biometrics 47, 139-146.

Korn E.L. (1982); Confidence Bands for Isotonic Dose-Response Curves, Appl. Statis. 31, 59-63.

Marcus R., Peritz E. and Gabriel, K.R. (1976); On closed testing procedures with special reference to ordered analysis of variance, Biometrika 63, 655-660.

Morgan, B.J.T. (1992); Analysis of Quantal Response Data, Chapman and Hall, London.

Mukerjee H., Robertson T. and Wright F. T. (1987); Comparison of several treatments with a control using multiple contrasts, Journal of the American Statistical Association, 82, 902 -910.

OECD (2006); Current approaches in the statistical analysis of ecotoxicity data: A guidance to application. Organisation for Economic Co-operation and Development. Report nr EVV/JM/MONO(2006)18 Number 54. 1-147.

Rao J.N.K. and Scott A.J. (1992) - A simple method for the analysis of clustered binary data, Biometrics 48, 577-585.

Rao J.N.K. and Scott A.J. (1999) - A simple method for analyzing overdispersion in clustered Poisson data, Statistics in Medicine 18, 1373-1385.

Robertson, T., Wright F.T. and Dykstra R.L. (1988); Order restricted statistical inference, Wiley.

Tarone, R.E. (1979); Testing the goodness of fit of the Binomial distribution, Biometrika 66, 585-590.

U.S.EPA (1995); The Use of the Benchmark Dose Approach in Health Risk Assessment, Risk Assessment Forum, EPA/630/R-94/007, United States Environmental Protection Agency, Washington, DC. Principal authors: K. Crump, B. Allen, E. Faustman.

Williams D.A. (1971); A test for differences between treatment means when several dose levels are compared with a zero dose control, Biometrics 27, 103-117.

Williams D.A. (1972); The comparison of several dose levels with a zero dose control, Biometrics 28, 519-531.

Williams D. A. (1975); The Analysis of Binary Responses from Toxicological Experiments Involving Reproduction and Teratotlogy, Biometrics 31, 949-952.

Williams D.A. (1977); Some inference procedures for monotonically ordered normal means, Biometrika 64, 9-14.

ANNEX 6 Statistical Protocol for Regression Estimates

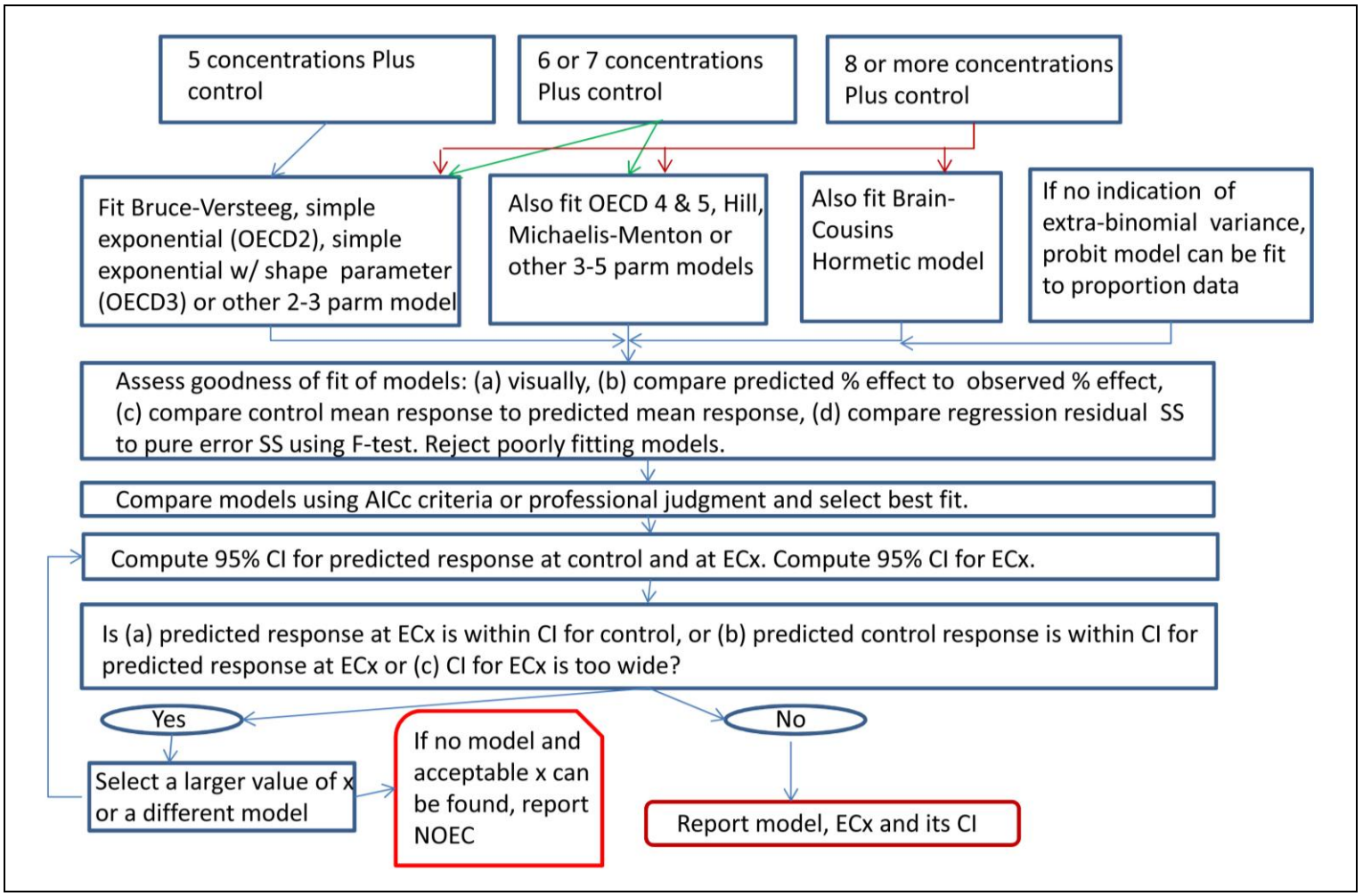


Figure 4: Flow chart for ECx Estimation of Replicate Mean Length, Weight, or Propartion of Egg Hatch or Larval Mortality

ANNEX 6 Statistical Protocol for Regression Estimates (continued)

General

The observations used to fit a model are replicate means (length and weight) or replicate proportions (egg hatch and larval mortality) (OECD 2006)

Weighted regression using replicate sample size as weight is generally advised. Other weighting schemes are possible, such as weighting by predicted mean response or a combination of this and replicate sample size. Weighting by reciprocal of within-concentration sample variance is not recommended (Bunke, Droge, Polzehl 1999, Seber and Wild 2003, Motulsky and Christopoulos 2004, Huet et al. (2003).

Any transformation of responses prior to analysis should preserve the independence of the observations and ECx and its confidence bounds should be expressed in the original units of measurement, rather than in transformed units. For example, a 20% change in the logarithm of length is not equivalent to a 20% change in length (Lyles et.al 2008, Draper and Smith 1999)

Considerations for Egg Hatch and Larval Mortality

For egg hatch and larval mortality, it is generally best to fit a decreasing model unless one is fitting a probit model as described below. That is, one should model the proportion of eggs that do not hatch or larvae that die. The reason for this is that ECx refers to a concentration at which there is a change equal to x% of the control mean response. If there are 5% control eggs that fail to hatch and one models failure to hatch, then EC20 refers to a concentration at which there is an change equal to 20% of the 5% control failure to hatch, and that is a change of 0.2*0.05=0.01 or 1 percentage point to 6% failure to hatch. Such a small change cannot be estimated in any meaningful way from the data available and is no interest in any case. Whereas if one models the proportion of eggs that hatch, the control proportion would be 95% in this example and a 20% reduction from the control mean would be a change of 0.95*0.2=0.18, so from 95% hatch success to 77% (= 95-18) hatching success and that effects concentration can be estimated and is presumably of greater interest. This is not an issue with size measurements, though adverse effects on size generally mean a decrease in size.

Models for Size (length or weight) and egg hatch success or larval survival. Except for the Brain-Cousins hermetic model, all of these models are described in Annex 7 and recommended in OECD (2006). What are called OECD 2-5, are also discussed for ecotoxicity experiments in Slob (2002). There are, of course, many other models that might be useful. Bunke, Droge, Polzehl (1999) lists numerous models not included here and references to other models are plentiful. Those listed below are suggested as particularly appropriate in ecotoxicity experiments and widely used.

<u>With 5 test concentrations plus control</u>

- Bruce-Versteeg
- Simple Exponential (OECD 2)
- Exponential with shape parameter (OECD 3)
- Simple Exponential with Lower Bound (OECD 4)

<u>With 6 or more test concentrations plus control</u>

- Exponential with shape parameter and lower bound (OECD 5)
- Michaelis-Menton
- Hill

Where there is visual evidence of hormesis (unlikely with egg hatch success or larval survival, but sometimes observed in size observations)

- Brain-Cousins Hormetic; Brain and Cousens (1989)

Alternative models for egg hatch failure and larval mortality

- Increasing models for these responses can be fit by probit (or logistic) models if there is no evidence of extra-binomial variance and control incidence is estimated in the model fit. This is not the preferred method, as it treats the individual, not the replicate, as the unit of analysis. (Morgan 1992, O'Hara Hines and Lawless 1993, Collett 2002, 2003)

Goodness of fit of a single model

- Visually compare observed and predicted percent decrease at each test concentration (Motulsky and Christopoulos 2004, Draper and Smith 1999).
- Compare regression error mean square against the pure error mean square using an F-test (Draper and Smith 1999).
- Check that every term in the model is significantly different from zero (i.e., determine whether all model terms are important), (Motulsky and Christopoulos 2004)
- Plots of residuals from regression vs. test concentration, possibly on a log(conc) scale. There should be no pattern to this plot; the points should be randomly scattered about a horizontal line at zero height.

Compare models

- Use Akiaki's AICc criteria. Smaller AICc values denote better fits and if AICc(B)-AICc(A)≥10, the model A is almost certainly better than model B. (Motulsky and Christopoulos (2004)
- Compare the two models visually by how well they meet the single model criteria above.
- The parsimony principal is advised, whereby the simplest model that fits the data reasonably well is used (Ratkowsky 1993, Lyles et.al 2008)

Quality of ECx estimate

The confidence interval (CI) for ECx should not be too wide. Statistical judgment is needed in deciding how wide the confidence interval can be and ECx still be useful. Simulations for regressions models fit to egg hatching and size data show that about 75% of confidence intervals for ECx (x=10, 20 or 30) span no more than two test concentrations. This provides a general guide for what is acceptable and a practical guide for what is achievable. Numerous authors assert the need to report confidence intervals for all model parameters and that wide confidence intervals for model parameters indicate unacceptable models (Ott & Longnecker 2008, Alvord and Rossio 1993, Motulsky and Christopoulos 2004, Lyles et.al 2008, Seber and Wild 2003, Bunke, Droge, Polzehl 1999, Environment Canada 2005).

The CI for ECx (or any other model parameter) should not contain zero Motulsky and Christopoulos (2004). Are the parameter estimates scientifically plausible? E.g., if the confidence interval for y0 is ±20%, no EC10 estimate is plausible. If the model predicts a 20% effect at a concentration C and the maximum observed effect at C and lower concentrations is 10%, then the EC20 is not plausible (Motulsky and Christopoulos 2004, Wang et al 2000, Environment Canada 2005)

ECx should not require extrapolation outside the range of positive concentrations Draper and Smith (1999), OECD (2006). For example, a general guide might be for ECx to be no more than about 25% below the lowest tested concentration or above the highest tested concentration.

The confidence interval for the effect at ECx should not contain zero. This is the regression equivalent the minimum significant difference that is often cited in hypothesis testing approaches (e.g., Wang et al 2000). It also corresponds to the confidence interval for the mean responses at the LOEC not contain the control mean.

References

Andersen, J. S., H. Holst, H. Spliid, H. Andersen, A. Baun and N. Nyholm (1998); Continuous Ecotoxicological Data

Evaluated Relative to a Control, Journal of Agricultural, Biological, and Environmental Statistics, Vol. 3, No. 4 (Dec.,1998), pp. 405-420

Alvord, W.G., Rossio, J.L. (1993); Determining confidence limits for drug potency in immunoassay, Journal of Immunological Methods 157, 155-163.

Bates, D. M., and D. G. Watts (1988); Nonlinear regression analysis and its applications, Wiley, New York.

Brain P. and Cousens R. (1989); An equation to describe dose responses where there is stimulation of growth at low doses. Weed res. 29: 93-96.

Bunke, O., Droge, B. and Polzehl, J. (1999). Model selection, transformations and variance estimation in nonlinear regression. *Statistics* 33, 197-240.

Collett, D. (2002); Modelling Binary Data, second edition, Chapman and Hall, London.

Collett, D. (2003); Modelling Survival Data in Medical Research, second edition, Chapman and Hall, London.

Draper, N.R. and Smith, H. (1999); Applied Regression Analysis, third edition. New York: John Wiley & Sons.

Environment Canada (2005); Guidance Document on Statistical Methods for Environmental Toxicity Tests, Report EPS 1/RM/46

Huet, S., A. Bouvier, M.-A. Poursat, E. Jolivet (2003); Statistical Tools for Nonlinear Regression: A Practical Guide with S-PLUS and R Examples, Springer Series in Statistics, New York.

Lyles, R. H., C. Poindexter, A. Evans, M. Brown, and C.R. Cooper (2008); Nonlinear Model-Based Estimates of IC50 for Studies Involving Continuous Therapeutic Dose-Response Data, *Contemp Clin Trials*. 2008 November ; 29(6): 878–886.

Morgan, B.J.T. (1992); Analysis of Quantal Response Data, Chapman and Hall, London.

Motulsky, H., A. Christopoulos (2004); Fitting Models to Biological Data Using Linear and Nonlinear Regression: A Practical Guide to Curve Fitting, Oxford University Press, USA.

O'Hara Hines, R. J. and J. F. Lawless (1993); Modelling Overdispersion in Toxicological Mortality Data Grouped over Time, *Biometrics* Vol. 49, pp. 107-121

OECD (2006); Current approaches in the statistical analysis of ecotoxicity data: A guidance to application. Organisation for Economic Co-operation and Development. Report nr EVV/JM/MONO(2006)18 Number 54. 1-147

Ott, R.L., M.T. Longnecker, An Introduction to Statistical Methods and Data Analysis, sixth edition, 2008, Brooks-Cole, Belmont, CA

Ratkowsky, D.A. (1993); Principles of nonlinear regression, Journal of Industrial Microbiology 12, 195-199.

Seber, G.A.F., C.J. Wild, Nonlinear Regression, Wiley, 2003

Slob W. (2002); Dose-response modelling of continuous endpoints. Toxicol. Sci., 66, 298-312

Wang, Q., D.L. Denton, and R. Shukla (2000); Applications and Statistical Properties Of Minimum Significant Difference-Based Criterion Testing In a Toxicity Testing Program, Environmental Toxicology and Chemistry, Vol. 19, pp. 113–117, 2000.